

2 KKKKKKKKKKKKKKKKKKKKK 21

Db 355 KKKKKKKKKKKKKKKKKKK 374

## RESULT 3

T49173

hypothetical protein T20N10.250 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 08-Dec-2000

C:Accession: T49173

R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; L

submitted to the Protein Sequence Database, April 2000

A:Reference number: 225017

A:Accession: T49173

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-517 &lt;DNA&gt;

A:Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250

A:Experimental source: cultivar Columbia; BAC clone T20N10

C:Genetics:

A:Gene: ATSP:T20N10.250

A:Map position: 3

A:Introns: 312/3: 359/3: 444/3

C:Superfamily: Arabidopsis thaliana hypothetical protein F17016.30

## Query Match

Best Local Similarity 91.7%; Score 100; DB 2; Length 517;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21

Db 444 KKKKKKKKKKKKKKKKKKK 463

## RESULT 4

T18513

hypothetical protein C0845c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000

C:Accession: T18513

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A:Reference number: 218935

A:Accession: T18513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-166 &lt;LAW&gt;

A:Cross-references: EMBL:298551; PIDN:CAB1123.2

C:Genetics:

A:Map position: 3

A:Introns: 19/1

A:Note: C0845c

## Query Match

Best Local Similarity 85.3%; Score 93; DB 2; Length 166;

Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21

Db 37 KKKKKKKKKKKKKKKKKKK 56

## RESULT 5

F71619

hypothetical protein PFB0235w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000

C:Accession: F71619

R:Gerdtner, M.J.; Tetelin, H.; Carucci, D.J.; Cummings, L.M.; Atavind, L.; Koonin, E.V.;

Perle, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 287, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743

A:Accession: F71619

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-483 &lt;GAR&gt;

A:Cross-references: GB:AE001382; GB:AE001362; NID:q3845130; PIDN:AC71836.1; PID:q384

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0235w

## Query Match

Best Local Similarity 82.6%; Score 90; DB 2; Length 483;

Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21

Db 449 KKKKKKKKKKKKKKKKKKK 468

## RESULT 6

C86477

protein F1504.29 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: C86477

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alor

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewart,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mailli, R.; Marzif

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talic

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C86477

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-107 &lt;STO&gt;

A:Cross-references: GB:AE005172; NID:98778346; PIDN:AAF79354.1; GSPDB:GN00141

C:Genetics:

A:Gene: F1504.29

A:Map position: 1

## Query Match

Best Local Similarity 79.8%; Score 87; DB 2; Length 107;

Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 19

Db 29 KKKKKKKKKKKKKKKKKKK 46

## RESULT 7

A48455

acidic phosphoprotein PCMA1g - Plasmodium chabaudi

C:Species: Plasmodium chabaudi

C:Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Mar-1998

C:Accession: A48455

R:Deleersnijder, W.; Prasomstli, P.; Tungpradubkul, S.; Hendrix, D.; Hamers-Casterm

Mol. Biochem. Parasitol. 56, 59-68, 1992

A:Title: Structure of a Plasmodium chabaudi acidic phosphoprotein that is associated

A:Reference number: A48455; MUID:93116806

A:Accession: A48455

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-441 &lt;DEL&gt;

A:Cross-references: GB:M95789; NID:g160602; PID:g160603

A:Experimental source: IP-PC1/C

A:Note: sequence extracted from NCBI backbone (NCBIN:121415, NCBI:P.121416)

C:Keywords: phosphoprotein

Query Match 77.1%; Score 84; DB 2; Length 441;  
Best Local Similarity 80.0%; Pred. No. 0.089;  
Matches 16; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKK 21  
|||||:|||||:|||||:  
Db 394 KKKKKKKKKKKKKKKKE 413

RESULT 8

T18440

hypothetical protein C0425v - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000

C:Accession: T18440

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A:Reference number: 218935

A:Accession: T18440

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-4550 <LAW>

A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CAE1121.1

C:Genetics:

A:Map position: 3

A:Note: C0425v

Query Match 74.3%; Score 81; DB 2; Length 4550;  
Best Local Similarity 44.2%; Pred. No. 0.77;  
Matches 19; Conservative 2; Mismatches 0; Indels 22; Gaps 1;

QY 1 CKK-----KKKKKKKKKKKKKK 21  
|||||:|||||:|||||:  
Db 132 CKKKMYFNKKKMYNEKFEYQNNKKKKKKKKKKKKKKKKKK 174

RESULT 9

T18452

hypothetical protein C0560C - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000

C:Accession: T18452

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, November 1998

A:Reference number: 218937

A:Accession: T18452

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-784 <LAW>

A:Cross-references: EMBL:AL008970; NID:e1407852; PID:e1332545; PIDN:CA1594.1

C:Genetics:

A:Map position: 3

A:Note: C0560C

Query Match 73.4%; Score 80; DB 2; Length 784;  
Best Local Similarity 75.0%; Pred. No. 0.31;  
Matches 15; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKK 21  
|||||:|||||:|||||:  
Db 360 EKKKKKKKKKKKKKKKK 379

RESULT 10

T50609

hypothetical protein DKFZP761B2423.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000

C:Accession: T50609  
R:Boecker, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Well, B.; Wiemann, S.  
submitted to the Protein Sequence Database, June 2000

A:Reference number: 225143

A:Accession: T50609

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-529 <AAA>

A:Cross-references: EMBL:AL359564

A:Experimental source: adult amygdala; clone DKFZP761B2423

C:Genetics:

A:Note: DKFZP761B2423.1

Query Match 71.6%; Score 78; DB 2; Length 529;  
Best Local Similarity 80.0%; Pred. No. 0.37;  
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKK 21  
|||||:|||||:|||||:  
Db 464 KKKKKKKKKKKKKKKKK 483

RESULT 11

T06377

SAR DNA-binding protein-1 - garden pea

C:Species: Pisum sativum (garden pea)

C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 02-Jun-2000

C:Accession: T06377

R:Hatton, D.; Gray, J.C.

submitted to the EMBL Data Library, April 1998

A:Description: cDNA encoding a pea SAR DNA-binding protein that shows homology to nuc

A:Reference number: 215637

A:Accession: T06377

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-560 <HAT>

A:Cross-references: EMBL:AF061962; NID:g3132695; PIDN:AAC16330.1; PID:g3132696

C:Genetics:

A:Gene: SARBP-1

C:Superfamily: garden pea SAR DNA-binding protein

Query Match 71.6%; Score 78; DB 2; Length 560;  
Best Local Similarity 75.0%; Pred. No. 0.38;  
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKK 21  
|||||:|||||:|||||:  
Db 463 KKKKKKKKKKKKKKKKK 482

RESULT 12

T42727

proliferation potential-related protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 02-Sep-2000

C:Accession: T42727

R:Witte, M.M.; Scott, R.E.

submitted to the EMBL Data Library, November 1998

A:Reference number: 22246

A:Accession: T42727

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1560 <WIT>

A:Cross-references: EMBL:U83913; NID:g385884; PID:g385885; PIDN:AAC72432.1

A:Experimental source: strain Balb/C

A:Gene: P2P-R

C:Function:

A:Description: involved in hnRNP association and Rb1 binding

C:Superfamily: RING finger homology

F:57-107/Domain: RING finger homology <RRN>







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:20:23 ; Search time 25.37 Seconds  
(without alignments)  
32.050 Million cell updates/sec

Title: US-09-461-684-1  
Perfect score: 109  
Sequence: 1 CXXXXXXXXXXXXXXXXXXXX 21

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt-40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	77.1	441	1	PHPA_PLACH
2	76	69.7	511	1	NOP5_YEAST
3	75	68.8	474	1	CBF5_SCHPO
4	75	68.8	504	1	SIR1_YEAST
5	75	68.8	2404	1	SON_MOUSE
6	75	68.8	2426	1	SON_HUMAN
7	74	67.9	414	1	Y694_METUA
8	74	67.9	474	1	CBF5_KLULA
9	74	67.9	483	1	CBF5_YEAST
10	74	67.9	726	1	BRD3_HUMAN
11	72	66.1	479	1	CBF5_CANAL
12	72	66.1	351	1	GARP_PLAF
13	71	65.1	686	1	CG79_HUMAN
14	71	65.1	686	1	CNG1_HUMAN
15	70.5	64.7	1411	1	TCOF_HUMAN
16	70	64.2	233	1	YJ08_YEAST
17	70	64.2	599	1	HM21_HUMAN
18	70	64.2	683	1	CNG1_RAT
19	70	64.2	684	1	CNG1_MOUSE
20	69.5	63.8	534	1	NOP5_YEAST
21	69.5	63.8	2231	1	SENI_YEAST
22	69	63.3	142	1	YMB8_YEAST
23	69	63.3	167	1	YK20_YEAST
24	69	63.3	723	1	SSRP_DROME
25	69	63.3	843	1	BLVR_BOVIN
26	68.5	62.8	724	1	Y061_GAEEL
27	68	62.4	523	1	DBP3_YEAST
28	68	62.4	1178	1	MNN4_YEAST
29	67	61.5	118	1	Y093_GAEEL
30	67	61.5	690	1	CNG1_BOVIN
31	67	61.5	691	1	CNG1_CANRA
32	67	61.5	1002	1	IF2P_YEAST
33	67	61.5	1220	1	IF2P_HUMAN

34	67	61.5	1362	1	BRD4_HUMAN	060885	homo sapien
35	66	60.6	481	1	CBF5_EMEHI	043100	emeritella
36	66	60.6	487	1	CBF5_ASFPU	043102	aspergillus
37	66	60.6	667	1	YE01_SCHPO	013796	schizosach
38	66	60.6	1153	1	A3D1_HUMAN	014617	homo sapien
39	66	60.6	1240	1	YNJ1_YEAST	P53935	saccharomyc
40	65.5	60.1	508	1	N060_DROME	044081	dirosophila
41	65.5	60.1	514	1	DKC1_HUMAN	060832	homo sapien
42	65	59.6	217	1	KS1_HYDAT	P38978	hydra atten
43	65	59.6	271	1	YGSW_YEAST	P53335	saccharomyc
44	65	59.6	320	1	YD33_YEAST	012117	saccharomyc
45	65	59.6	344	1	YR02_YEAST	P38079	saccharomyc

## ALIGNMENTS

```
RESULT 1
ID PHPA_PLACH STANDARD; PRT; 441 AA.
AC 002752;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 01-JUN-1994 (Rel. 29, Last annotation update)
DE Acidic phosphoprotein precursor (50 kDa antigen).
GN PCMA1.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5825;
RN 11;
RP SEQUENCE FROM N.A.
RC STRAIN=IP-PC1;
RX MEDLINE=93116806; PubMed=1475002;
RA Deleersnijder W., Prasomsitti P., Tungradubkul S., Hendrix D.,
RA Hamers-Casterman C., Hamers R.;
RT "Structure of a Plasmodium chabaudi acidic phosphoprotein that is
RT associated with the host erythrocyte membrane."
RL Mol. Biochem. Parasitol. 56:59-68(1992).
CC -1- FUNCTION: DURING INFECTION, THIS PHOSPHOPROTEIN PROBABLY MODULATES
CC THE STRUCTURE OF THE RED CELL MEMBRANE TO THE ADVANTAGE OF THE
CC PARASITE, ALTHOUGH ITS PRECISE FUNCTION IS NOT KNOWN.
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ON THE
CC CYTOPLASMIC FACE OF THE HOST ERYTHROCYTE MEMBRANE.
CC -1- MISCELLANEOUS: ASSOCIATED WITH THE HOST RED CELL MEMBRANE
CC THROUGHOUT THE ENTIRE ERYTHROCYTIC CYCLE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, M5789; AAA29732.1; -
DR PIR, A48455; A48455.
KW Phosphorylation; Signal; Antigen; Membrane; Repeat; Erythrocyte.
FT SIGNAL 1 15
FT CHAIN 16 441
FT DOMAIN 186 313
FT REPEAT 186 193
FT REPEAT 194 201
FT REPEAT 202 209
FT REPEAT 210 217
FT REPEAT 218 225
FT REPEAT 226 233
FT REPEAT 234 241
FT REPEAT 242 249
FT REPEAT 250 257
FT REPEAT 258 265
FT REPEAT 266 273
FT REPEAT 274 281
FT REPEAT 282 289
FT 1-13.
```

FT REPEAT 290 297 1-14.  
FT REPEAT 298 305 1-15.  
FT REPEAT 306 313 1-16.  
FT DOMAIN 353 370 2 X 9 AA TANDEM REPEATS.  
FT REPEAT 353 360 2-1.  
FT REPEAT 361 368 2-2.  
FT DOMAIN 371 417 LYS-RICH (BASIC).  
FT CARBOHD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 441 AA: 49708 MW: DB85E83E795EE7E5 CRC64;

Query Match 77.1% Score 84; DB 1; Length 441;  
Best Local Similarity 80.0%; Pred. No. 0.023;  
Matches 16; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Oy 2 KKKKKKKKKKKKKKKKK 21  
DB 394 KKKKKKKKKKKKKKK 413

RESULT 2  
NOP5\_YEAST  
ID NOP5\_YEAST STANDARD: PRT: 511 AA.  
AC 012499;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nucleolar protein NOP58 (Nucleolar protein NOP5).  
GN NOP58 OR NOP5 OR YOR10C OR O6108.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / FY1679;  
RA Pearson B.M., Hernandez Y., Wolf S.S., Kalogeropoulos A., Schweizer M.;  
RL Submitted (Aug-1993) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RC STRAIN-W303;  
RA MEDLINE=98298165; PubMed=9632712;  
RX Wu P., Brockenhrough J.S., Metcalfe A.C., Chen S., Arts J.P.;  
RT "NOP5p is a small nucleolar ribonucleoprotein component required for  
pre-18S rRNA processing in yeast."  
RT J. Biol. Chem. 273:16453-16463(1998).  
CC -1- FUNCTION: REQUIRED FOR PRE-18S RNA PROCESSING. MAY BIND  
MICROTUBULES.  
CC -1- SUBUNIT: INTERACTS WITH NOP56 AND NOP1.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.  
CC -1- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X90565; CAA62165.1; -;  
DR EMBL: 275217; CAA96630.1; -;  
DR EMBL: AF056070; AAC39484.1; -;  
DR SCD: S0005837; NOP58.  
DR InterPro: IPR002687; NOP.  
DR Pfam: PF01798; NOP.1.  
DR ProDom: PD004104; NOP.1.  
KW Ribosome biogenesis; Nuclear protein; rRNA processing.  
FT DOMAIN 441 511  
SQ SEQUENCE 511 AA: 56956 MW: 8A288944B2A1E2 CRC64;

Query Match 69.7% Score 76; DB 1; Length 511;  
Best Local Similarity 70.0%; Pred. No. 0.15;  
Matches 14; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Oy 2 KKKKKKKKKKKKKKKKK 21  
DB 480 KKKKKKKKKKKKKKK 499

RESULT 3  
CBF5\_SCHPO  
ID CBF5\_SCHPO STANDARD: PRT: 474 AA.  
AC 014007;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Centromere/microtubule binding protein cbf5 (Centromere-binding factor  
5) (Nucleolar protein cbf5).  
GN CBF5 OR SPAC29A4.04C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;  
RA Brown D., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;  
RL Submitted (Jul-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A  
CENTROMERE DNA-CBF5-BINDING FACTOR AND IS INVOLVED IN MITOTIC  
CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH (BY  
SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (by similarity).  
CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: 297210; CAB10131.1; -;  
DR InterPro: IPR002478; PUA.  
DR InterPro: IPR002501; TRUB\_N.  
DR Pfam: PF01472; PUA.1.  
DR Pfam: PF01509; TRUB\_N.1.  
DR SMART: SM00359; PUA.1.  
KW Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.  
FT DOMAIN 434 468  
FT REPEAT 443 445 1.  
FT REPEAT 450 452 2.  
FT REPEAT 454 456 3.  
FT REPEAT 457 459 4.  
FT REPEAT 460 462 5.  
FT REPEAT 463 465 6.  
FT REPEAT 466 468 7.  
SQ SEQUENCE 474 AA: 53110 MW: B8C9896C5FAEB71 CRC64;

Query Match 68.8% Score 75; DB 1; Length 474;  
Best Local Similarity 73.7%; Pred. No. 0.18;  
Matches 14; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Oy 2 KKKKKKKKKKKKKKKKK 20  
DB 454 KKKKKKKKKKKKKKK 472

RESULT 4

SIKL\_YEAST  
ID SIKL\_YEAST STANDARD: PRT: 504 AA.  
AC 012460:  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE SIKL protein (Nucleolar protein NOP56).  
OS SIKL OR NOP56 OR YLR197W OR L8167.9.  
OC Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TPH1;  
RX MEDLINE=96040178; PubMed=7547500;  
RA Morin P.J., Downs J.A., Snodgrass A.M., Gilmore T.D.;  
RT "Genetic analysis of growth inhibition by GAL4-L kappa B-alpha in  
RT Saccharomyces cerevisiae.";  
RL Cell Growth Differ. 6:789-798(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
RA Favell A., Fulton L., Galtung S., Greco T., Kirsten J., Kucaba T.,  
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,  
RA Johnston L., Langston Y., Latreille P., Mardis E., Meneses S.,  
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,  
RA Tach S., Trevasakis E., Vaudin M., Vignati D., Wilcox L., Wilson R.,  
RA Wohldman P., Waterston R.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP CHARACTERIZATION, AND MUTAGENESIS.  
RX MEDLINE=98038777; PubMed=9372940;  
RA Gautier T., Berges T., Tollervey D., Hurt E.;  
RT "Nucleolar KKE/D repeat proteins NOP56p and NOP58p interact with Nop1p  
RT and are required for ribosome biogenesis.";  
RL Mol. Cell. Biol. 17:7088-7098(1997).  
CC -1- FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT SYNTHESIS.  
CC -1- SUBUNIT: INTERACTS WITH NOP1 AND NOP58.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR; CYTOSOL.  
CC -1- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.  
CC CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: U20237; AAC49066.1; -  
DR EMBL: U14913; AAB67431.1; -  
DR SCD: S0004187; SIK1.  
DR InterPro: IPR002687; NOP.  
DR Pfam: PF01798; NOP.1.  
DR Prodom: PD004104; NOP.1.  
KM Ribosome biogenesis; Nuclear  
FT DOMAIN 443 504  
FT 443 504 ASP/GLU/LYS-RICH.  
FT MUTAGEN 333 333  
FT 333 333 V->A: REDUCED GROWTH RATE AT ALL  
FT 333 333 TEMPERATURES; WHEN ASSOCIATED WITH R-385.  
FT MUTAGEN 355 355 Y->C: AT 37 DEGREES, GROWTH SLOWS AFTER  
FT 355 355 NO 8 HOURS AND CELL DIVISION STOPS AFTER  
FT 355 355 20 HOURS.  
FT MUTAGEN 385 385 M->R: REDUCED GROWTH RATE AT ALL  
FT 385 385 TEMPERATURES; WHEN ASSOCIATED WITH A-313.  
FT SEQUENCE 504 AA; 56864 MW; F8522A5870EF4842 CRC64;  
Query Match 68.8%; Score 75; DB 1; Length 504;  
Best Local Similarity 70.0%; Pred. No. 0.19;  
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
Db 465 KKKKKKKKKKKKKKKKK 484  
RESULT 5  
ID SON\_MOUSE STANDARD: PRT: 2404 AA.  
AC 09QX47; 09QX45; 09QX06; 09QX12;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE SON protein.  
GN SON.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC STRAIN=129/SV;  
RX MEDLINE=204008866; PubMed=10950926;  
RA Wynn S.L., Fisher R.A., Pagel C., Price W., Liu Q.Y., Khan I.M.,  
RA Zammit P., Dadgar K., Mazzanti W., Kessling A., Lee J.S., Bulwela L.;  
RT "Organization and conservation of the GART/SON/DONSON locus in mouse  
RT and human genomes.";  
RL Genomics 68:57-62(2000).  
RN [2]  
RP SEQUENCE OF 1-116 FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Hippocampus, Small intestine, and Tongue;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanake I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochila H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsch G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -1- FUNCTION: Transcriptional repressor. Binds to the consensus DNA  
CC sequence: 5'-GATCT[ATG][AG]CC-3'. Might protect cells from  
CC apoptosis. Might be involved in pre-mRNA splicing (By similarity).  
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
CC produced by alternative splicing.  
CC -1- TISSUE SPECIFICITY: Widely expressed.  
CC -1- DOMAIN: Contains 8 types of repeats which are distributed in 3  
CC regions.  
CC -1- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.  
CC CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AF193606; AAF23120.1; -  
DR EMBL: AF193595; AAF23120.1; JOINED.



RA Chumakov I.M., Berdichevskii F.B., Sokolova N.V., Reznikov M.V.,  
RA Prasolov V.S.;  
RT "Identification of a protein product of a novel human gene SON and  
RT the biological effect upon administering a changed form of this gene  
RT into mammalian cells.";  
RT Mol. Biol. (Mosk) 25:731-740(1991).  
RN [8]  
RN SEQUENCE OF 709-1079 FROM N.A. (ISOFORM I).  
RP TISSUE-Placenta;  
RC MEDLINE=93062885; PubMed=1435774;  
RX Bliskovskii V.V., Kirillov A.V., Zakhariev V.M., Chumakov I.M.;  
RT "The human son gene: the large and small transcripts contains various  
RT 5'-terminal sequences";  
RT Mol. Biol. (Mosk) 26:807-812(1992).  
RN [9]  
RN SEQUENCE OF 1009-1131 FROM N.A.  
RP TISSUE-Placenta;  
RC MEDLINE=93062884; PubMed=1435773;  
RX Bliskovskii V.V., Berdichevskii F.B., Tkachenko A.V., Belova M.E.,  
RA Chumakov I.M.;  
RT "Coding part of the son gene small transcript contains four areas of  
RT complete tandem repeats.";  
RT Mol. Biol. (Mosk) 26:793-806(1992).  
RN [10]  
RN SEQUENCE OF 1145-2426 FROM N.A. (ISOFORM F).  
RP MEDLINE=93048367; PubMed=1424986;  
RX Mactoul T., Hume C.R., Konigorski S., Hayes P., Osterweil Z.,  
RA Lee J.S.;  
RT "A cDNA clone for a novel nuclear protein with DNA binding  
RT activity.";  
RT Chromosoma 101:618-624(1992).  
RN [11]  
RN SEQUENCE OF 1692-2175 FROM N.A. (ISOFORM A).  
RP MEDLINE=89039788; PubMed=3054499;  
RX Berdichevskii F.B., Chumakov I.M., Kiselev L.L.;  
RT "Decoding of the primary structure of the son3 region in human  
RT genome: identification of a new protein with unusual structure and  
RT homology with DNA-binding proteins.";  
RT Mol. Biol. (Mosk) 22:794-801(1988).  
RN [12]  
RN SEQUENCE OF 1939-2426 FROM N.A. (ISOFORM J).  
RP TISSUE-Cerebellum;  
RX MEDLINE=99439804; PubMed=10509013;  
RA Greenhalf W., Lee J., Chaudhuri B.;  
RT "A selection system for human apoptosis inhibitors using yeast.";  
RT Yeast 15:1107-1132(1999).  
RN [13]  
RN -I- FUNCTION: Represses hepatitis B virus (HBV) core promoter activity  
RN and transcription of HBV genes and production of HBV virions.  
RN Binds to the consensus DNA sequence: 5'-GAG[GT]AN[CG][AG]CC-3'.  
RN Might protect cells from apoptosis. Might be involved in pre-mRNA  
RN splicing (by similarity).  
RN -I- SUBCELLULAR LOCATION: Nuclear with a speckled distribution.  
RN -I- ALTERNATIVE PRODUCTS: 10 isoforms; A, B, C, D, E, F (shown here),  
RN G, H, I and J; may be produced by alternative splicing.  
RN -I- TISSUE SPECIFICITY: Widely expressed, with the higher expression  
RN seen in leukocyte and heart.  
RN -I- DOMAIN: contains 8 types of repeats which are distributed in 3  
RN regions.  
RN -I- MISCELLANEOUS: Colocalizes with the pre-mRNA splicing factor  
RN SFRS2/SC-35.  
RN -I- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.  
RN -I- SIMILARITY: CONTAINS 1 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.  
RN -I- CAUTION: ISOFORM A SEQUENCE FROM REF.7 DIFFERS FROM THAT SHOWN  
RN DUE TO A FRAMESHIFT.  
RN -I- CAUTION: ISOFORM F SEQUENCE FROM REF.10 DIFFERS FROM THAT SHOWN  
RN DUE TO A FRAMESHIFT.  
RN -----  
RN This SWISS-PROT entry is copyright. It is produced through a collaboration  
RN between the Swiss Institute of Bioinformatics and the EMBL outstation-  
RN the European Bioinformatics Institute. There are no restrictions on its  
RN use by non-profit institutions as long as its content is in no way  
RN modified and this statement is not removed. Usage by and for commercial  
RN entities requires a license agreement (See <http://www.isb-slb.ch/announce/>)

CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch).  
-----  
CC EMBL: AF380179; AAL34497.1; -  
DR EMBL: X63753; CAA45282.1; ALT\_FRAME.  
DR EMBL: M36428; AAA36624.1; -  
DR EMBL: AF380180; AAL34498.1; -  
DR EMBL: AF380181; AAL34499.1; -  
DR EMBL: AF380182; AAL34500.1; -  
DR EMBL: AF380183; AAL34501.1; -  
DR EMBL: AF380184; AAL34502.1; -  
DR EMBL: AY026895; AAK07692.1; -  
DR EMBL: AF435977; AAL30810.1; -  
DR EMBL: X63751; CAC69885.1; -  
DR EMBL: AB028942; BAA82971.1; -  
DR EMBL: X63071; CAA44793.1; ALT\_FRAME.  
DR EMBL: AF139897; AAD50078.1; -  
DR EMBL: SA72238; AAB23945.1; -  
DR EMBL: AK024752; BAB14985.1; -  
DR EMBL: AF161428; AAF28988.1; -  
DR EMBL: AF161430; AAF28990.1; -  
DR PIR: PNO099; PNO099.  
DR MIM: 182465; -  
DR InterPro: IPR001159; DS\_RBD.  
DR InterPro: IPR000467; G\_patch.  
DR Pfam: PF00035; dsrm; 1.  
DR Pfam: PF01585; G\_patch; 1.  
DR SMART: SM00358; DSRM; 1.  
DR SMART: SM00443; G\_patch; 1.  
DR PROSITE: PS50137; DS\_RBD; 1.  
DR PROSITE: PS50174; G\_PATCH; 1.  
KW RNA-binding; DNA-binding; Nuclear protein; Repeat;  
KW Alternative splicing.  
FT DOMAIN 726 895  
FT FT  
FT 912 988  
FT FT  
FT DOMAIN  
FT FT  
FT 1006 1126  
FT REPEAT 1006 1011  
FT REPEAT 1014 1019  
FT REPEAT 1021 1026  
FT REPEAT 1030 1035  
FT REPEAT 1038 1043  
FT REPEAT 1046 1051  
FT REPEAT 1055 1060  
FT REPEAT 1063 1068  
FT REPEAT 1071 1076  
FT REPEAT 1080 1085  
FT REPEAT 1089 1094  
FT REPEAT 1100 1105  
FT REPEAT 1111 1116  
FT REPEAT 1121 1126  
FT DOMAIN 1147 1179  
FT FT  
FT 1359 1390  
FT FT  
FT DOMAIN  
FT FT  
FT 1925 1994  
FT REPEAT 1925 1931  
FT REPEAT 1953 1959  
FT REPEAT 1960 1966  
FT FT

Query Match 68.8%; Score 75; DB 1; Length 2426;  
Best Local Similarity 75.0%; Pred. No. 0.61;  
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
DB 109 KKKKKKKKKKKKKKKKKKK 128

RESULT 7  
Y694\_METJA  
ID Y694\_METJA STANDARD; PRT; 414 AA.  
AC Q58105;

```
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M00694.
GN M00694.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
NC NCBL_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bull C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake O., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.J.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii."
RL Science 273:1058-1073(1996).
CC -i- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67516; AAB98689.1; -
DR TIGR: M00694; -
DR InterPro: IPR002687; NOP.
DR Pfam: PF01798; NOP. 1.
DR ProDom: PD004104; NOP. 1.
KW Hypothetical protein; Complete proteome.
FT DOMAIN 349 414 ASP/GLU/LYS-RICH.
SJ SEQUENCE 414 AA; 47799 MW; A9092EFC3C82C407 CRC64;
```

```
Query Match 67.9%; Score 74; DB 1; Length 414;
Best Local Similarity 75.0%; Pred. No. 0.2;
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```
OY 2 KKKKKKKKKKKKKKKKKKK 21
DB 377 KKKKKKKKKKKKKKKKK 396

RESULT 8
CBF5_KLUJLA STANDARD: PRT; 474 AA.
AC 013473;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Centromere/microtubule binding protein CBF5 (Centromere-binding factor
DE 5) (Nucleolar protein CBF5).
GN CBF5.
OS Kluyveromyces fragilis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
NC NCBL_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBD100;
RX MEDLINE=98144788; PubMed=9483794;
RA Winkler A.A., Bobok A., Zonneveld B.J.M., Steensma H.Y.,
RA Hooykaas P.J.J.;
```

```
RT "The lysine-rich C-terminal repeats of the centromere-binding factor
RT 5 (CBF5) of Kluyveromyces fragilis are not essential for function."
RL Yeast 14:37-48(1998).
CC -i- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
CC SOME WAY ASSOCIATED WITH THE CBF3 110 KDA SUBUNIT (CBF3A) (BY
CC SIMILARITY).
CC -i- SUBCELLULAR LOCATION: Nuclear; nucleolar (by similarity).
CC -i- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF008563; AAC64862.1; -
DR InterPro: IPR002478; PUA.
DR Pfam: PF01472; PUA. 1.
DR SMART: SM00359; PUA. 1.
KW Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
FT DOMAIN 431 460
FT REPEAT 431 433 1.
FT REPEAT 434 436 2.
FT REPEAT 437 439 3.
FT REPEAT 440 442 4.
FT REPEAT 443 445 5.
FT REPEAT 446 448 6.
FT REPEAT 449 451 7.
FT REPEAT 452 454 8.
FT REPEAT 455 457 9.
SJO SEQUENCE 474 AA; 53630 MW; 95306CE7FEA756C CRC64;
```

```
Query Match 67.9%; Score 74; DB 1; Length 474;
Best Local Similarity 70.0%; Pred. No. 0.22;
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 2 KKKKKKKKKKKKKKKKKKK 21
DB 434 KKKKKKKKKKKKKKKKK 453

RESULT 9
CBF5_YEAST STANDARD: PRT; 483 AA.
AC P33372;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Centromere/microtubule binding protein CBF5 (Centromere-binding factor
DE 5) (Nucleolar protein CBF5) (P64').
GN CBF5 OR YNR175W OR L9470.11.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NC NCBL_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=93330283; PubMed=8336724;
RA Jiang W., Middleton K., Yoon H.-J., Fouquet C., Carbon J.;
RA "An essential yeast protein, CBF5p, binds in vitro to centromeres and
RA microtubules."
RL Mol. Cell. Biol. 13:4884-4893(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
```

```

RA Favella A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kacaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Miles L., Taich A., Treviski E., Vignati D.,
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.:
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
CC SOME WAY ASSOCIATED WITH THE CBF3 110 KDA SUBUNIT (CBF3A).
CC -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR.
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L12351; AAA34473.1; -.
DR EMBL: U17246; AAB67463.1; -.
DR PIR: S41853; S41853.
DR SGD: S0004165; CBF5.
DR InterPro: IPR002478; CBF5.
DR InterPro: IPR002501; TRUB_N.
DR Pfam: PF01472; PUA: 1.
DR Pfam: PF01509; TRUB_N: 1.
DR SMART: SM00359; PUA: 1.
DR Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding
KW DOMAIN 434 463 10 X 3 AA TANDEM REPEATS OF K-K-IDEI.
FT REPEAT 437 436 1.
FT REPEAT 437 439 2.
FT REPEAT 440 442 3.
FT REPEAT 443 445 4.
FT REPEAT 446 448 5.
FT REPEAT 449 451 6.
FT REPEAT 452 454 7.
FT REPEAT 455 457 8.
FT REPEAT 458 460 9.
FT REPEAT 461 463 10.
SQ SEQUENCE 483 AA; 54704 MW; D356B39FDC32E2D CRC64;

Query Match 67.9%; Score 74; DB 1; Length 483;
Best Local Similarity 70.0%; Pred. No. 0.23;
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKK 21
Db 434 KKKKKKKKKKKKKKKKK 453

RESULT 10
BRD3_HUMAN STANDARD: PRT: 726 AA.
AC 015059: 092645;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Bromodomain-containing protein 3 (RING3-like protein).
GN BRD3 OR RING3L OR KIAA0043.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,

```

```

RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.:
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (K1AA0041-K1AA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1."
RL DNA Res. 1:223-229(1994).
RN [2]
RP SEQUENCE OF 363-726 FROM N.A.
RX MEDLINE=98038990; PubMed=9373153;
RA Thorpe K.L., Gorman P., Thomas C., Sheer D., Trowsdale J., Beck S.:
RT "Chromosomal localization, gene structure and transcription pattern of
RT the ORFX gene, a homologue of the MHC-linked RING3 gene."
RL Gene 200:177-183(1997).
CC -!- SUBCELLULAR LOCATION: Nuclear (potential).
CC -!- TISSUE SPECIFICITY: UBQUITOUS.
CC -!- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D26362; BAA05393.1; -.
DR EMBL: Z81330; CAB03630.1; -.
DR MIM: 601541; -.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain: 2.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO: 2.
DR PROSITE: PS00633; BROMODOMAIN_1; 2.
DR PROSITE: PS50014; BROMODOMAIN_2; 2.
KW Bromodomain; Repeat; Nuclear protein.
FT DOMAIN 56 115 BROMODOMAIN 1.
FT DOMAIN 326 398 BROMODOMAIN 2.
FT DOMAIN 487 555 LYS-RICH.
FT DOMAIN 676 725 SER-RICH.
FT CONFLICT 465 466 EL -> DV (IN REF. 2).
SQ SEQUENCE 726 AA; 79541 MW; 64F526FC3C1033AA CRC64;

Query Match 67.9%; Score 74; DB 1; Length 726;
Best Local Similarity 70.0%; Pred. No. 0.31;
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKK 21
Db 489 KKKKKKKKKKKKKKKKK 508

RESULT 11
CBF5_CANAL STANDARD: PRT: 479 AA.
AC 043101;
DT 15-JUL-1998 (Rel. 36; Created)
DT 15-JUL-1998 (Rel. 36; Last sequence update)
DT 15-DEC-1998 (Rel. 37; Last annotation update)
DE Centromere/microtubule binding protein CBF5 (Centromere-binding factor
DE 5) (Nucleolar protein CBF5).
GN CBF5.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RA Jiang W., Clifford J., Koltin Y.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
CC SOME WAY ASSOCIATED WITH THE CBF3 110 KDA SUBUNIT (CBF3A) (BX

```



```
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U59149; AAB94297.1; -
CC DR InterPro: IPR002478; PUA.
CC DR InterPro: IPR002501; TRUB_N.
CC DR Pfam: PF01472; PUA.1.
CC DR Pfam: PF01509; TRUB_N.1.
CC SMART: SM00359; PUA.1.
CC Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
CC SEQUENCE 479 AA; 54321 MW; 3BAF5104E12C9EB6 CRC64;

Query Match 66.1%; Score 72; DB 1; Length 479;
Best Local Similarity 70.0%; Pred. No. 0.35;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
   ||| ||| ||| ||| ||| |||
Db 432 KKKKKKKKKKKKKKKKKKK 451

RESULT 12
GARP_PLAUF STANDARD; PRT; 678 AA.
ID GARP_PLAUF
AC P13816;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutamic acid-rich protein precursor.
GN GARP.
OS Plasmodium falciparum (Isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89040048; PubMed=2903445;
RA Triglia T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RA Kemp D.J.;
RT *Structure of a Plasmodium falciparum gene that encodes a glutamic
RT acid-rich protein (GARP).";
RL MOL. Biochem. Parasitol. 31:199-202(1988).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J03998; AAA29605.1; -
CC DR PIR: A54514; A54514.
CC Repeat: Malaria; Antigen; Signal.
CC SIGNAL 1 25
FT CHAIN 26 678 GLUTAMIC ACID-RICH PROTEIN.
FT FT 120 164 15 X 3 AA TANDEN REPEATS OF K-K-X.
FT DOMAIN 372 416 9 X APPROXIMATE TANDEN REPEATS.
FT DOMAIN 417 441 5 X APPROXIMATE TANDEN REPEATS.
FT DOMAIN 576 604 POLY-GLU.
FT DOMAIN 605 653 7 X APPROXIMATE TANDEN REPEATS.
FT DOMAIN 654 663 POLY-GLU.
FT SEQUENCE 678 AA; 80551 MW; 2A6F85606496E9E CRC64;
```

```
Query Match 66.1%; Score 72; DB 1; Length 678;
Best Local Similarity 70.0%; Pred. No. 0.46;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
   ||| ||| ||| ||| ||| |||
Db 132 KKKKKKKKKKKKKKKKKKK 151

RESULT 13
CG79_HUMAN STANDARD; PRT; 351 AA.
ID CG79_HUMAN
AC O9Y388;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CGI-79.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20272150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT Identification of novel human genes evolutionarily conserved in
RT Caenorhabditis elegans by comparative proteomics.";
RL Genome Res. 10:703-713(2000).
CC
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF151837; AAD34074.1; -
CC DR HSSP: P09651; 1HA1.
CC DR InterPro: IPR000504; RRM.
CC DR Pfam: PF00076; rrm.1.
CC DR SMART: SM00360; RRM; 1.
CC DR PROSITE: PS50102; RRM; 1.
CC DR PROSITE: PS00030; RRM_RNP.1; 1.
CC KW Hypothetical protein; RNA-binding.
CC FT DOMAIN 36 114 RNA-BINDING (RRM).
CC SEQUENCE 351 AA; 39675 MW; 7B6E882D6B192EBE CRC64;

Query Match 65.1%; Score 71; DB 1; Length 351;
Best Local Similarity 73.7%; Pred. No. 0.35;
Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 20
   ||| ||| ||| ||| ||| |||
Db 156 KKKKKKKKKKKKKKKKKKK 174

RESULT 14
CNG1_HUMAN STANDARD; PRT; 686 AA.
ID CNG1_HUMAN
AC P29973; Q16485; Q16279;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE cGMP-gated cation channel alpha 1 (CNG channel alpha 1) (CNGC1)
DE (Cyclic nucleotide-gated channel alpha 1) (Cyclic nucleotide-gated
DE channel, photoreceptor) (Cyclic-nucleotide-gated cation channel 1)
DE (Rod photoreceptor cGMP-gated channel alpha subunit).
GN CNGA1 OR CNCG1 OR CNCG.
OS Homo sapiens (Human).;
```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=92210603; PubMed=1372902;  
 RA Plettner S.J., Lee A.K., Altherr M.R., Howard T.A., Seldin M.F.,  
 RA Hurvitz R.L., Wasmuth J.J., Baehr W.;  
 RT "Primary structure and chromosomal localization of human and mouse  
 rod photoreceptor cGMP-gated cation channel.";  
 RL J. Biol. Chem. 267:6257-6262(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=92356211; PubMed=1379636;  
 RA Dhallan R.S., Macke J.P., Eddy R.L., Shows T.B., Reed R.R.,  
 RA Yau K.-W., Nathans J.;  
 RT "Human rod photoreceptor cGMP-gated channel: amino acid sequence,  
 RT gene structure, and functional expression.";  
 RL J. Neurosci. 12:3248-3256(1992).  
 RN [3]  
 RP SEQUENCE OF 313-573 FROM N.A.  
 RX MEDLINE=95175019; PubMed=7532814;  
 RA Distler M., Biel M., Flocke V., Hofmann F.;  
 RT "Expression of cyclic nucleotide-gated cation channels in non-sensory  
 RT tissues and cells.";  
 RL Neuropharmacology 33:1275-1282(1994).  
 RN [4]  
 RP VARIANT ARRP PHE-316, AND VARIANTS GLN-28 AND ASN-114.  
 RX MEDLINE=96036047; PubMed=7479749;  
 RA Dryja T.P., Finn J.T., Peng Y.-W., McGee T.L., Barton E.L., Yau K.-W.;  
 RT "Mutations in the gene encoding the alpha subunit of the rod  
 RT cGMP-gated channel in autosomal recessive retinitis pigmentosa.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:10177-10181(1995).  
 CC -1- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN  
 CC COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN  
 CC BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION  
 CC CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD  
 CC PHOTORECEPTORS.  
 CC -1- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER. FORMS HETEROOLIGOMERIC  
 CC COMPLEX WITH CNGB4.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: ROD CELLS IN THE RETINA.  
 CC -1- DISEASE: DEFECTS IN CNGB4 ARE A CAUSE OF AUTOSOMAL RECESSIVE  
 CC RETINITIS PIGMENTOSA (ARRP). A DISEASE THAT LEADS TO DEGENERATION  
 CC OF RETINAL PHOTORECEPTOR CELLS.  
 CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL  
 CC FAMILY.  
 CC -1- DATABASE: NAME=Mutations of the CNGB4 gene;  
 CC NOTE=Retina International's Scientific Newsletter;  
 CC WWW="http://www.retina-international.com/sci\_news/cngalmut.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: M84741; AAA52010.1; ALT\_INIT.  
 DR EMBL: S42457; AAB22778.1; -  
 DR EMBL: S76052; AAD14206.1; -  
 DR PIR: A42161; A42161.  
 DR MIM: 123825; -  
 DR InterPro: IPR000636; Catlon\_chan\_non\_11g.  
 DR InterPro: IPR000595; CNMP\_binding.  
 DR Pfam: PF00027; CNMP\_binding. 1.  
 DR Pfam: PF00520; Ion\_trans. 1.  
 DR SMART: SM00100; CNMP. 1.  
 DR PROSITE: PS00888; CNMP\_BINDING\_1; 1.  
 DR PROSITE: PS00889; CNMP\_BINDING\_2; 1.

DR PROSITE: PS0042; CNMP\_BINDING\_3. 1.  
 KW Ionic channel; Ion transport; CAMP-binding; Transmembrane;  
 KW Multigene family; Vision; Disease mutation; Polymorphism;  
 KW Retinitis pigmentosa.  
 FT DOMAIN 1 160  
 FT TRANSMEM 161 181  
 FT DOMAIN 182 194  
 FT TRANSMEM 195 213  
 FT DOMAIN 214 237  
 FT TRANSMEM 238 257  
 FT DOMAIN 258 295  
 FT TRANSMEM 296 318  
 FT DOMAIN 319 370  
 FT TRANSMEM 371 390  
 FT DOMAIN 391 474  
 FT TRANSMEM 475 495  
 FT DOMAIN 496 686  
 FT NP\_BIND 483 605  
 FT BINDING 542 542  
 FT BINDING 557 557  
 FT CARBOHYD 421 421  
 FT VARIANT 28 28  
 FT /Frid=VAR\_009295.  
 FT VARIANT 114 114  
 FT /Frid=VAR\_009296.  
 FT VARIANT 316 316  
 FT CONFLICT 46 46  
 FT CONFLICT 85 85  
 FT CONFLICT 146 147  
 FT CONFLICT 539 539  
 FT CONFLICT 677 678  
 FT SEQUENCE 686 AA; 79126 MW; B5200D216FC97AF6 CRC64;  
 QY 2 KKKKKKKKKKKKKKKKKKK 21  
 Db 124 KKKKKKKKKKKKKKKKKKK 143  
 Query Match 65.1%; Score 71; DB 1; Length 686;  
 Best Local Similarity 70.0%; Pred. No. 0.57;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 RESULT 15  
 ID TCOF\_HUMAN STANDARD; PRT; 1411 AA.  
 AC 013428; 099408;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-Oct-2001 (Rel. 40, Last annotation update)  
 DE Treacher protein (Treacher collins syndrome protein).  
 GN TCOF1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96154183; PubMed=8563749;  
 RA Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.K.,  
 RA Bonner C.A., Koprivnikar K., Wasmuth J.J.;  
 RT "Positional cloning of a gene involved in the pathogenesis of  
 RT Treacher Collins syndrome. The Treacher Collins Syndrome  
 RT Collaborative Group.";  
 RL Nat. Genet. 12:130-136(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97250498; PubMed=9096354;  
 RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,  
 RA Ashley J.A., Lovett M., Jabs E.W.;  
 RT "TCOF1 gene encodes a putative nucleolar phosphoprotein that exhibits  
 RT mutations in Treacher Collins syndrome throughout its coding

RT region.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).  
RN [3]  
RP VARIANTS L-439; V-810; V-1313 AND G-1355, AND VARIANT TCS R-53.  
RX MEDLINE-97195537; PubMed-9042910;  
RA Edwards S.J., Gladwin A.J., Dixon M.J.:  
RT "The mutational spectrum in Treacher Collins syndrome reveals a  
RT predominance of mutations that create a premature-termination  
RT codon."  
RL Am. J. Hum. Genet. 60:515-524(1997).  
CC -1- DISEASE: DEFECTS IN TCOF1 ARE THE CAUSE OF TREACHER COLLINS  
CC SYNDROME (TCS). TCS IS A AUTOSOMAL DOMINANT DISORDER OF  
CC CRANIOFACIAL DEVELOPMENT THAT OCCURS WITH AN INCIDENCE OF 1/50,000  
CC LIVE BIRTHS. THE CLINICAL FEATURES OF TCS ARE BILATERALLY  
CC SYMMETRICAL AND INCLUDE: (1) ABNORMALITIES OF THE EXTERNAL EARS,  
CC ATRESIA OF THE EXTERNAL EAR CANALS, AND MALFORMATION OF THE MIDDLE  
CC EAR OSSICLES, WHICH MAY RESULT IN CONDUCTIVE HEARING LOSS. (2)  
CC LATERAL DOWNGARD SLOPING OF PALPEBRAL FISSURES, FREQUENTLY WITH  
CC COLOBOMAS OF THE LOWER EYELIDS; (3) HYPOPLASIA OF THE MANDIBLE AND  
CC ZYGOMATIC COMPLEX; (4) CLEFT PALATE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U40847; AAC50903.1; -  
DR EMBL: U76366; AAC51181.1; -  
DR EMBL: U84664; AAC51185.1; -  
DR EMBL: U84640; AAC51185.1; JOINED.  
DR EMBL: U84641; AAC51185.1; JOINED.  
DR EMBL: U84642; AAC51185.1; JOINED.  
DR EMBL: U84643; AAC51185.1; JOINED.  
DR EMBL: U84644; AAC51185.1; JOINED.  
DR EMBL: U84645; AAC51185.1; JOINED.  
DR EMBL: U84646; AAC51185.1; JOINED.  
DR EMBL: U84647; AAC51185.1; JOINED.  
DR EMBL: U84648; AAC51185.1; JOINED.  
DR EMBL: U84649; AAC51185.1; JOINED.  
DR EMBL: U84650; AAC51185.1; JOINED.  
DR EMBL: U84651; AAC51185.1; JOINED.  
DR EMBL: U84652; AAC51185.1; JOINED.  
DR EMBL: U84653; AAC51185.1; JOINED.  
DR EMBL: U84654; AAC51185.1; JOINED.  
DR EMBL: U84655; AAC51185.1; JOINED.  
DR EMBL: U84656; AAC51185.1; JOINED.  
DR EMBL: U84657; AAC51185.1; JOINED.  
DR EMBL: U84658; AAC51185.1; JOINED.  
DR EMBL: U84659; AAC51185.1; JOINED.  
DR EMBL: U84660; AAC51185.1; JOINED.  
DR EMBL: U84661; AAC51185.1; JOINED.  
DR EMBL: U84662; AAC51185.1; JOINED.  
DR EMBL: U84663; AAC51185.1; JOINED.  
DR MIM: 154500; -  
DR InterPro: IPR003993; treacle.  
DR PRINTS: PR01503; TREACLE.  
KW Disease mutation: Polymorphism.  
FT DOMAIN 89 97  
FT DOMAIN 204 207  
FT DOMAIN 616 619  
FT DOMAIN 919 924  
FT DOMAIN 1285 1289  
FT DOMAIN 1375 1386  
FT DOMAIN 1398 1405  
FT VARIANT 53 53  
FT P -> R (IN TCS).  
FT /FTID-VAR\_005630.  
FT P -> L.  
FT /FTID-VAR\_005631.  
FT A -> V.  
FT /FTID-VAR\_005632.

FT VARIANT 1313 1313 A -> V.  
FT /FTID-VAR\_005633.  
FT VARIANT 1355 1355 D -> G.  
FT /FTID-VAR\_005634.  
SO SEQUENCE 1411 AA: 144312 MW: 3880203D985C2699 CRC64:

Query Match 64.7%; Score 70.5; DB 1; Length 1411;  
Best Local Similarity 54.8%; Pred. No. 1.1; Indels 11; Gaps 1;  
Matches 17; Conservative 2; Mismatches 1;

QY 2 KKKKKKKKKKK-----KKKKKKK 21  
I:|||||  
Db 1375 KKKKKKKKKAKASTKDSPPSKKKKKKK 1405

Search completed: July 1, 2002, 06:31:33  
Job time: 670 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:28 ; Search time 83.98 Seconds  
(without alignments)  
43.259 Million cell updates/sec

Title: US-09-461-684-1

Perfect score: 109  
Sequence: 1 CXXXXXXXXXXXXXXXXXXXX 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mhc:\*  
8: SP-organelle:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-virus:\*  
13: SP-vertebrate:\*  
14: SP-unclassified:\*  
15: SP-virus:\*  
16: SP-bacteriophage:\*  
17: SP-archaeoprotein:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	91.7	128	3	09P529
2	100	91.7	129	11	035807
3	100	91.7	128	4	09H5V6
4	100	91.7	215	11	064075
5	100	91.7	260	10	09L629
6	100	91.7	380	10	09NT34
7	100	91.7	517	10	09LXR2
8	100	91.7	531	6	0951V6
9	100	91.7	720	4	09H607
10	97	89.0	667	4	09H607
11	90	82.6	257	4	09H607
12	90	82.6	483	5	09H5Y3
13	90	82.6	686	4	09NFX0
14	89	81.7	83	11	09ER82
15	87	79.8	107	10	09LOP6
16	81	74.3	4550	5	077336

17	80	73.4	130	4	09H7B1	09H7B1 homo sapien
18	80	73.4	213	11	P97762	P97762 mus musculu
19	80	73.4	784	5	077339	077339 plasmidum
20	79	72.5	260	11	099LX5	099LX5 mus musculu
21	78	71.6	263	4	09H76	09H76 homo sapien
22	78	71.6	407	4	000536	000536 homo sapien
23	78	71.6	529	4	09NPX4	09NPX4 homo sapien
24	78	71.6	560	10	065334	065334 plasm saliv
25	78	71.6	628	4	09H5M5	09H5M5 homo sapien
26	78	71.6	1591	11	P97868	P97868 mus musculu
27	78	71.6	1616	4	096PH3	096PH3 homo sapien
28	77.5	71.1	353	4	09H8Q9	09H8Q9 homo sapien
29	77	70.6	103	5	09NM87	09NM87 trypanosoma
30	77	70.6	229	11	09JM93	09JM93 mus musculu
31	77	70.6	266	10	09X134	09X134 arabidopsis
32	77	70.6	402	4	09H407	09H407 homo sapien
33	77	70.6	577	3	09C1M8	09C1M8 schizosacch
34	77	70.6	749	5	0967D9	0967D9 drosophila
35	77	70.6	3724	5	077320	077320 plasmidum
36	76	69.7	581	5	096549	096549 dicystostei
37	76	69.7	812	5	021827	021827 caenorhabd
38	76	69.7	1629	5	09U0K9	09U0K9 plasmidum
39	75	68.8	2086	11	09OXP5	09OXP5 mus musculu
40	75	68.8	2404	11	09OX47	09OX47 mus musculu
41	74	67.9	191	5	045682	045682 caenorhabd
42	74	67.9	198	16	097E51	097E51 clostridium
43	74	67.9	244	11	09CWL7	09CWL7 mus musculu
44	74	67.9	256	11	09CUI4	09CUI4 mus musculu
45	74	67.9	550	5	09VVF1	09VVF1 drosophila

#### ALIGNMENTS

RESULT 1  
ID 09P529 PRELIMINARY: PRT: 128 AA.  
AC 09P529.  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DT HYPOTHETICAL 15.2 KDA PROTEIN.  
GN B24H17.160.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AL356815; CAB92638.2; --  
KW Hypothetical protein.  
SQ SEQUENCE 128 AA: 15157 MW; 8C7C65C3DFB70765 CRC64;

Query Match 91.7%: Score 100; DB 3; Length 128;  
Best Local Similarity 100.0%; Pred. No. 9.6e-05;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KXXXXXXXXXXXXXXXXXXXX 21  
DB 71 KXXXXXXXXXXXXXXXXXXXX 90

RESULT 2  
ID 035807 PRELIMINARY: PRT: 129 AA.

AC 035607;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE MICROVASCULAR ENDOTHELIAL DIFFERENTIATION PROTEIN 2.  
 GN MDC2.  
 OS Rattus norvegicus (rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-EPIDIDYMIS;  
 RX MEDLINE=96172708; PubMed=9511718;  
 RA Proels F., Loser B., Marx M.;  
 RT "Differential expression of osteopontin, PC4, and CEC5, a novel mRNA  
 RT species, during in vitro angiogenesis.";  
 RL Exp. Cell Res. 239:1-10(1998).  
 DR EMBL: Y08769; CAZ70022.1;  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR Pfam: PF00069; Pkinase.1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM; 1.  
 KW ATP-binding; Transferase.  
 SQ SEQUENCE 129 AA; 15080 MW; 38102272BBE2EDB4 CRC64;

Query Match 91.7%; Score 100; DB 11; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 9,7e-05;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 85 KKKKKKKKKKKKKKKKKKK 104

RESULT 3  
 ID 09H5V6 PRELIMINARY; PRT; 168 AA.  
 AC 09H5V6;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE CDNA: FLJ22976 FIS, CLONE KAT11222 (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
 RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,  
 RA Nakamura Y., Isono T., Sugano S.;  
 RT "NDO human cDNA sequencing project.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK026629; BAB15513.1;  
 FT NON-TER 168  
 SQ SEQUENCE 168 AA; 19549 MW; A19DBD195F8A1A90 CRC64;

Query Match 91.7%; Score 100; DB 4; Length 168;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 140 KKKKKKKKKKKKKKKKKKK 159

RESULT 4  
 ID 064075 PRELIMINARY; PRT; 215 AA.  
 AC 064075;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE NUCLEOPORIN P62 HOMOLOG PROTEIN (FRAGMENT).  
 OS Rattus sp.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10118;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95151924; PubMed=7649178;  
 RA Wang Z.Q., Akmal K.M., Kim K.H.;  
 RT "An unusual nucleoporin-related messenger ribonucleic acid is present  
 RT in the germ cells of rat testis.";  
 RL Biol. Reprod. 51:1022-1030(1994).  
 DR EMBL: S75997; AAB33384.1;  
 KW Porin.  
 FT NON-TER 1  
 SQ SEQUENCE 215 AA; 24593 MW; 098251C97A8FBD88 CRC64;

Query Match 91.7%; Score 100; DB 11; Length 215;  
 Best Local Similarity 100.0%; Pred. No. 0.00014;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 35 KKKKKKKKKKKKKKKKKKK 54

RESULT 5  
 ID 09LG29 PRELIMINARY; PRT; 260 AA.  
 AC 09LG29;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE:FL109.  
 OS Arabidopsis thaliana (mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana Chromosome 3. III.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP002460; BAA97098.1;  
 DR InterPro: IPR001386; Linker\_histone.  
 DR PRINTS: PR00624; HISTONEH5.  
 SQ SEQUENCE 260 AA; 33307 MW; 43E2394CB8131143 CRC64;

Query Match 91.7%; Score 100; DB 10; Length 260;  
 Best Local Similarity 100.0%; Pred. No. 0.00016;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 7 KKKKKKKKKKKKKKKKKKK 26

RESULT 6  
 ID 09NT34 PRELIMINARY; PRT; 380 AA.  
 AC 09NT34;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HYPOTHEETICAL 42.7 KDA PROTEIN (FRAGMENT).  
 GN DKP2P43411120.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Oltjenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL137556; CAB70810.1; -  
 KW Hypothetical protein.  
 FT NON-TER 380  
 SQ SEQUENCE 380 AA; 42689 MW; 6750DD101346AFB CRC64;

Query Match 91.7%; Score 100; DB 4; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 0.00025;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKK 21  
 DB 355 KKKKKKKKKKKKKKK 374

RESULT 7  
 O9LXR2 PRELIMINARY; PRT; 517 AA.  
 AC O9LXR2;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DE 01-OCT-2000 (TREMBlrel. 18, Last annotation update)  
 DE HYPOTHETICAL 59.7 KDA PROTEIN.  
 GN 20N10.250.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,  
 RA Rued S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL353032; CAB88307.1; -  
 DR InterPro: IPR001810; F-box.  
 DR Pfam: PF00646; F-box; 1.  
 DR SMART: SM00256; FBOX; 1.  
 DR PROSITE: PS50181; FBOX; 1.  
 KW Hypothetical protein  
 SQ SEQUENCE 517 AA; 59689 MW; EC6D957D01F86E70 CRC64;

Query Match 91.7%; Score 100; DB 10; Length 517;  
 Best Local Similarity 100.0%; Pred. No. 0.00025;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKK 21  
 DB 444 KKKKKKKKKKKKKKK 463

RESULT 8  
 O9SLV6 PRELIMINARY; PRT; 531 AA.  
 AC O9SLV6;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOTHETICAL 61.4 KDA PROTEIN (FRAGMENT).  
 OX Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 RN [1]

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
 OC Cercopitheciinae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,  
 RA Terao K., Sugano S.;  
 RT Isolation of novel full-length cDNA clones from macaque testis cDNA  
 RT libraries."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB071085; BAB64479.1; -  
 KW Hypothetical protein.  
 FT NON-TER 531  
 SQ SEQUENCE 531 AA; 61389 MW; B55996B4F5CDD60C CRC64;

Query Match 91.7%; Score 100; DB 6; Length 531;  
 Best Local Similarity 100.0%; Pred. No. 0.00025;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKK 21  
 DB 502 KKKKKKKKKKKKKKK 521

RESULT 9  
 O9H607 PRELIMINARY; PRT; 720 AA.  
 AC O9H607;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE cDNA: FLJ21979 FIS, CLONE HEP06065 (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,  
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
 RT NEDO human cDNA sequencing project."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK025632; BAB15196.1; -  
 FT NON-TER 720  
 SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;

Query Match 91.7%; Score 100; DB 4; Length 720;  
 Best Local Similarity 100.0%; Pred. No. 0.00031;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKK 21  
 DB 692 KKKKKKKKKKKKKKK 711

RESULT 10  
 O9HC48 PRELIMINARY; PRT; 667 AA.  
 AC O9HC48;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE CTCL TUDOR ANTIGEN SE2-5 (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE-TESTIS;  
RX MEDLINE-21143360; PubMed-11149944;  
RA Eichmuller S., Usener D., Dummer R., Stein A., Thiel D.,  
Schadendorf D.;  
RT "Serological detection of cutaneous T-cell lymphoma-associated  
antigens."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:629-634(2001).  
DR EMBL: AF177228; AAG33676.1; -  
DR HSSP: 012923; 3PDZ.  
DR InterPro: IPR001478; PDZ.  
DR Pfam: PF00595; PDZ; 2.  
DR SMART: SM00228; PDZ; 2.  
DR PROSITE: PS50106; PDZ; 2.  
FT NON\_TER 1  
FT 667 667  
SQ SEQUENCE 667 AA; 73499 MW; C653EC16802BAE02 CRC64;

Query Match 89.0%; Score 97; DB 4; Length 667;  
Best Local Similarity 95.0%; Pred. No. 0.00063;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
DB 648 KKKKKKKKKKKKKKKKK 667

RESULT 11  
O9H5Y3  
ID 09H5Y3 PRELIMINARY: PRT; 257 AA.  
AC 09H5Y3;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE CDNA: FLJ22789 FIS, CLONE KAI2171 (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-ILEAL MUCOSA;  
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,  
Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,  
Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,  
RA Isogai T., Sugano S.;  
RT "NEO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AK026442; BAB15484.1; -  
DR NON\_TER 257  
FT 257  
SQ SEQUENCE 257 AA; 28778 MW; 100DEA6ABA521868 CRC64;

Query Match 82.6%; Score 90; DB 4; Length 257;  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 19  
DB 240 KKKKKKKKKKKKKKKKK 257

RESULT 12  
O96148  
ID 096148 PRELIMINARY: PRT; 483 AA.  
AC 096148;  
DT 01-MAY-1999 (TReMBLrel. 10, Created)  
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)  
DE HYPOTHEICAL 57.8 KDA PROTEIN.  
GN PFB0235W.

OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-99021743; PubMed-9804551;  
RA Gardner M.J., Tetteh H., Carucci D.J., Cummings L.M., Aravind L.,  
Rao Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,  
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perera M.,  
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium  
falciparum."  
RL Science 282:1126-1132(1998).  
DR EMBL: AE001382; AAC71836.1; -  
DR Hypothetical protein  
SQ SEQUENCE 483 AA; 57785 MW; 9332C1EC9B4DECFCB CRC64;

Query Match 82.6%; Score 90; DB 5; Length 483;  
Best Local Similarity 85.0%; Pred. No. 0.0029;  
Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
DB 449 KKKKKKKKKKKKKKKKK 468

RESULT 13  
O9NXXF0  
ID 09NXXF0 PRELIMINARY: PRT; 686 AA.  
AC 09NXXF0;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE CDNA FLJ20288 FIS, CLONE HEP0414 (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,  
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
RT "NEO human cDNA sequencing project."  
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AK000295; BAA91063.1; -  
DR HSSP: Q00420; IAWC.  
DR InterPro: IPR002110; ANK.  
DR Pfam: PF00023; ANK; 14.  
DR SMART: SM00248; ANK; 14.  
DR PROSITE: PS50088; ANK\_REPEAT; 13.  
DR PROSITE: PS50297; ANK\_REPEAT; 1.  
DR ANK repeat; Repeat.  
FT NON\_TER 686  
FT 686  
SQ SEQUENCE 686 AA; 73615 MW; AE1380212E0FAD22 CRC64;

Query Match 82.6%; Score 90; DB 4; Length 686;  
Best Local Similarity 85.0%; Pred. No. 0.0037;  
Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
DB 666 KKKKKKKKKKKKKKKKK 685

RESULT 14  
O9ER82  
ID 09ER82 PRELIMINARY: PRT; 83 AA.  
AC 09ER82;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)



Search completed: July 1, 2002, 06:30:54  
Job time: 686 sec

DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE SR528 PROTEIN (FRAGMENT).  
CN SR528.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/10; TISSUE=SKLETAL MUSCLE;  
RA Kemp T.J., Sadusky T.J., Carey N., Coulton G.R.;  
RT "Murine genes identified from skeletal muscle mRNA which had undergone  
RT 7 days of passive stretch in vivo."  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
FR EMBL: A250693; CAC08505.1; -.  
FT NON\_TER  
SQ SEQUENCE 83 AA: 10093 MW: 3088B2113D02DE92 CRC64;

Query Match 81.7%; Score 89; DB 11; Length 83;  
Best Local Similarity 85.0%; Pred. No. 0.0011;  
Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
Db 49 KKKKKKKKKKKKKKKKKKK 68

RESULT 15  
O9LOF6 PRELIMINARY; PRT: 107 AA.  
AC O9LOF6;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE F1504.29.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chlou J., Choi E.,  
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsy N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Thayerl A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC007887; AAF9354.1; -.  
SQ SEQUENCE 107 AA: 13427 MW: 91091726AED96615 CRC64;

Query Match 79.8%; Score 87; DB 10; Length 107;  
Best Local Similarity 94.4%; Pred. No. 0.0022;  
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 19  
Db 29 KKKKKKKKKKKKKKKKKKK 46



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 1, 2002, 06:10:53 ; Search time 98.97 Seconds

(Without alignments)  
23.568 Million cell updates/sec

Title: US-09-461-684-1

Perfect score: 109

Sequence: 1 CRRKKKKKKKKKKKKKKKKKK 21

## Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.GeneSeq\_032802.\*

```
1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.*
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109	100.0	21	AA13780	Soluble peptide an
2	109	100.0	39	AAW45801	One chain of a bom
3	109	100.0	58	AAU18238	Novel human DNA-bi
4	109	100.0	70	AAO11210	Human polypeptide
5	109	100.0	81	AAO03766	Human polypeptide
6	105	96.3	36	AAO11849	Human polypeptide
7	105	96.3	47	AAO12447	Human polypeptide
8	105	96.3	66	AAU18239	Novel human DNA-bi
9	105	96.3	69	AAU18239	Human colon cancer
10	105	96.3	83	AAO37729	Human colon cancer
11	103	94.5	42	AAO09258	Human polypeptide

12	103	94.5	68	AAO11214	Human polypeptide
13	103	94.5	124	AAO00291	Human polypeptide
14	102	93.6	41	AAO12203	Human polypeptide
15	102	93.6	60	AAO12187	Human polypeptide
16	102	93.6	62	AAO12187	Human polypeptide
17	102	93.6	69	AAO12187	Human polypeptide
18	102	93.6	69	AAO12187	Human polypeptide
19	102	93.6	108	AAU18184	Novel human DNA-bi
20	102	93.6	108	AAU18184	Human polypeptide
21	101	92.7	54	AAO00092	Novel human DNA-bi
22	101	92.7	54	AAO00092	Novel human DNA-bi
23	101	92.7	99	AAO07806	Human polypeptide
24	100	91.7	20	AAO20159	Sequence of lysine
25	100	91.7	22	AAW38833	Delivery peptide u
26	100	91.7	22	AAW38833	Delivery peptide u
27	100	91.7	23	AAW38833	Delivery peptide u
28	100	91.7	23	AAW38833	Delivery peptide u
29	100	91.7	23	AAW38833	Delivery peptide u
30	100	91.7	23	AAW38833	Delivery peptide u
31	100	91.7	24	AAW38833	Delivery peptide u
32	100	91.7	24	AAW38833	Delivery peptide u
33	100	91.7	25	AAW38833	Delivery peptide u
34	100	91.7	25	AAW38833	Delivery peptide u
35	100	91.7	25	AAW38833	Delivery peptide u
36	100	91.7	25	AAW38833	Delivery peptide u
37	100	91.7	26	AAW38833	Delivery peptide u
38	100	91.7	26	AAW38833	Delivery peptide u
39	100	91.7	26	AAW38833	Delivery peptide u
40	100	91.7	26	AAW38833	Delivery peptide u
41	100	91.7	26	AAW38833	Delivery peptide u
42	100	91.7	27	AAW38833	Delivery peptide u
43	100	91.7	27	AAW38833	Delivery peptide u
44	100	91.7	27	AAW38833	Delivery peptide u
45	100	91.7	27	AAW38833	Delivery peptide u

## ALIGNMENTS

RESULT 1  
AAB13780 standard; peptide: 21 AA.  
XX  
AC AAB13780;  
XX  
DT 10-NOV-2000 (first entry)  
XX  
DE Soluble peptide antigen pk.  
XX  
KW pk peptide: cytotoxic T cell; CTL; immunotherapy;  
KW major histocompatibility complex class I; MHC class I; antigen; tumour;  
KW prostate; breast; multiple myeloma.  
XX  
OS Unidentified.  
XX  
PN WO200035949-A1.  
XX  
PD 22-JUN-2000.  
XX  
PF 14-DEC-1999; 99WO-US29724.  
XX  
PR 14-DEC-1999; 98US-0112324.  
XX  
PA (DEND-) DENDREON CORP.  
XX  
PI Laus R, Hakim I, Vldovic D;  
XX  
DR WPI: 2000-442365/38.  
XX  
PT Antigens modified by the covalent addition of a peptide that  
PT facilitates entry into antigen presenting cells, useful for producing  
PT compositions for immunizing against tumors and pathogens -



PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225216.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0228287.  
PR 01-SEP-2000; 2000US-0228287.  
PR 01-SEP-2000; 2000US-0228343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0233968.  
PR 14-SEP-2000; 2000US-0233997.  
PR 14-SEP-2000; 2000US-0233998.  
PR 14-SEP-2000; 2000US-0233999.  
PR 14-SEP-2000; 2000US-0234000.  
PR 14-SEP-2000; 2000US-0234201.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 23-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI: 2001-465557/50.  
XX N-PSDB: AAS29114.  
XX  
PT Nucleic acid molecules encoding human secreted chromosomal binding  
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
PT Alzheimer's and Parkinson's diseases and cancers -  
XX  
XX  
PS Claim 11; SEQ ID No 223; 561pp; English.  
XX  
XX  
CC The present invention relates to the isolation of novel DNA-binding  
CC proteins, and cDNA (AAS29030-AAS29157) and genomic sequences encoding  
CC for these proteins. DNA-binding proteins such as histones, chromo  
CC (chromatin organisation modifier) domain proteins, and Y-box binding  
CC proteins may contribute to diseases resulting from aberrant DNA  
CC organisation and/or gene transcription. The sequences of the invention  
CC are useful in screening assays to identify antagonists and/or agonists  
CC that may enhance or block activities mediated by DNA-binding proteins.  
CC Blockers of DNA-binding proteins may be useful in treating disorders  
CC such as malignant diseases (e.g. cancer), autoimmune disorders  
CC (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid  
CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious  
CC diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's

CC disease). The polynucleotide sequences of the invention may also be  
CC used in gene therapy. AA018154-AA018281 represent novel DNA-binding  
CC proteins.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 58 AA:

Query Match 100.0%; Score 109; DB 22; Length 58;  
Best Local Similarity 100.0%; Pred. No. 5e-06;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CXXXXXXXXXXXXXXXXX 21  
|XXXXXXXXXXXXXXXX|  
DB 35 cXXXXXXXXXXXXXXXXX 55

RESULT 4

AA011210  
ID AA011210 standard; Protein; 70 AA.

XX AC AA011210;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 25102.

XX KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;

XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

XX KM tissue growth factor; immunomodulatory; cancer; leukaemia;

XX KM nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YF, Liu C, Drmanac RT;

XX DR WPI; 2001-514838/56.

XX DR N-PSDB; AA191141.

XX PT Isolated nucleic acids and polypeptides, useful for preventing  
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune  
XX PT disorders -

XX PS Claim 20; SEQ ID NO 25102; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 70 AA:

Query Match 100.0%; Score 109; DB 22; Length 70;  
Best Local Similarity 100.0%; Pred. No. 5.8e-06;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CXXXXXXXXXXXXXXXXX 21  
|XXXXXXXXXXXXXXXX|  
DB 30 cXXXXXXXXXXXXXXXXX 50

RESULT 5

AA003766  
ID AA003766 standard; Protein; 81 AA.

XX AC AA003766;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 17658.

XX KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;

XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

XX KM tissue growth factor; immunomodulatory; cancer; leukaemia;

XX KM nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YF, Liu C, Drmanac RT;

XX DR WPI; 2001-514838/56.

XX DR N-PSDB; AA183697.

XX PT Isolated nucleic acids and polypeptides, useful for preventing  
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune  
XX PT disorders -

XX PS Claim 20; SEQ ID NO 17658; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 81 AA:

Query Match 100.0%; Score 109; DB 22; Length 81;  
Best Local Similarity 100.0%; Pred. No. 6.4e-06;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CXXXXXXXXXXXXXXXXX 21  
DB 48 CXXXXXXXXXXXXXXXXX 68

## RESULT 6

AA01849  
ID AA01849 standard; Protein: 36 AA.

AC AA01849;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 25741.

OS Human: cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KM tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001: 2001WO-US04927.

PR 28-FEB-2000: 2000US-0515126.

PR 18-MAY-2000: 2000US-0577409.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-514838/56.

DR N-PSDB: AA191780.

PT Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -

PS Claim 20: SEQ ID NO 25741: 1399pp + Sequence Listing: English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and

CC the encoded proteins (AA000010-AA03910) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 36 AA:

Query Match 96.3%; Score 105; DB 22: Length 36;

Best Local Similarity 95.2%; Pred. No. 9.6e-06;

Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CXXXXXXXXXXXXXXXXX 21

DB 11 CXXXXXXXXXXXXXXXXX 31

## RESULT 7

AA012447

ID AA012447 standard; Protein: 47 AA.

AC AA012447;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 26339.

OS Human: cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KM tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001: 2001WO-US04927.

PR 28-FEB-2000: 2000US-0515126.

PR 18-MAY-2000: 2000US-0577409.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-514838/56.

DR N-PSDB: AA192378.

PT Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -

PS Claim 20: SEQ ID NO 26339: 1399pp + Sequence Listing: English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and

CC the encoded proteins (AA000010-AA03910) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 47 AA:

Query Match 96.3%; Score 105; DB 22: Length 47;

Best Local Similarity 95.2%; Pred. No. 1.2e-05;

Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CXXXXXXXXXXXXXXXXX 21

DB 12 CXXXXXXXXXXXXXXXXX 32

## RESULT 8

AA018239

AC AA018239;

DT 21-NOV-2001 (first entry)

DE Novel human DNA-binding protein #86.

XX Human: DNA-binding protein; histone; chromo domain protein;  
KW Chromatin organisation modifier; Y-box binding protein;  
KW DNA organisation; gene transcription; malignant disease;  
KW autoimmune disorder; rheumatic disease; genetic abnormality;  
KW infectious disease; neurological disorder; gene therapy;  
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;  
KW Cytostatic.  
XX  
OS Homo sapiens.  
XX  
PN MO200155162-A1.  
PD  
XX 02-AUG-2001.  
PF  
XX 17-JAN-2001; 2001MO-US01305.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225477.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226682.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0233397.  
PR 14-SEP-2000; 2000US-0233398.  
PR 14-SEP-2000; 2000US-0233399.  
PR 14-SEP-2000; 2000US-0233400.

PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234297.  
PR 25-SEP-2000; 2000US-0234498.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239835.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.





XX (HUMA-) HUMAN GENOME SCI INC.  
PA Ruben SM, Barash SC, Birse CE, Rosen CA:  
XX WPI: 2001-235357/24.  
XX N-PSDB: AAH33160.  
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers -  
PS Claim 11: Page 6294: 9803pp: English.  
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the activity of P by producing of P.  
CC Additionally, N may be used to produce the colon cancer-associated P,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAB77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX Sequence 83 AA:  
SQ  
Query Match 96.3%; Score 105; DB 22; Length 83;  
Best Local Similarity 95.2%; Pred. No. 1.8e-05;  
Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CRRKKKKKKKKKKKKKKKK 21  
I:|||||  
Db 47 cekkkskkkkkkkkkkkkkk 67  
RESULT 11  
AAO09258  
ID AAO09258 standard; Protein: 42 AA.  
XX  
AC AAO09258;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 23150.  
XX  
KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
XX  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX

PI Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI: 2001-514838/56.  
DR N-PSDB: AAI91189.  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20: SEQ ID NO 23150; 1399pp + Sequence Listing: English.  
XX  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activity/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 42 AA:  
SQ  
Query Match 94.5%; Score 103; DB 22; Length 42;  
Best Local Similarity 95.2%; Pred. No. 1.8e-05;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CRRKKKKKKKKKKKKKKKK 21  
I:|||||  
Db 5 cpkkskkkkkkkkkkkkkk 25  
RESULT 12  
AAO11214  
ID AAO11214 standard; Protein: 68 AA.  
XX  
AC AAO11214;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 25106.  
XX  
KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
XX  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI: 2001-514838/56.  
DR N-PSDB: AAI91145.  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -  
XX  
PS Claim 20: SEQ ID NO 25106; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 68 AA:  
  
Query Match 94.5%; Score 103; DB 22; Length 68;  
Best Local Similarity 95.2%; Pred. No. 2.6e-05;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 CXXXXXXXXXXXXXXXXX 21  
|XXXXXXXXXXXXXXXXX  
Db 17 CXXXXXXXXXXXXXXXXX 37  
  
RESULT 13  
AA000291  
ID AA000291 standard; Protein: 124 AA.  
XX  
AC AA000291:  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 14183.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI: 2001-514838/56.  
DR N-PSDB: AA180222.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20: SEQ ID NO 14183; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 124 AA:  
  
Query Match 94.5%; Score 103; DB 22; Length 124;  
Best Local Similarity 95.2%; Pred. No. 4.2e-05;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 CXXXXXXXXXXXXXXXXX 21  
|XXXXXXXXXXXXXXXXX  
Db 20 CXXXXXXXXXXXXXXXXX 40  
  
RESULT 14  
AA012203  
ID AA012203 standard; Protein: 41 AA.  
XX  
AC AA012203:  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 26095.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI: 2001-514838/56.  
DR N-PSDB: AA192134.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20: SEQ ID NO 26095; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukemia, nervous system disorders, arthritis and

CC Inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 41 AA:

Query Match 93.6%: Score 102; DB 22: Length 41;  
Best Local Similarity 95.2%: Pred. No. 2.3e-05;  
Matches 20: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CXXXXXXXXXXXXXXXXX 21  
1 |XXXXXXXXXXXXXXXX|  
DB 4 cXXXXXXXXXXXXXXXXX 24

RESULT 15

AAO12187  
ID AAO12187 standard; Protein: 60 AA.

XX AAO12187:

XX 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 26079.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

XX nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

XX 07-SEP-2001.

PD 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-514838/56.

DR N-PSDB; AA192118.

XX Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -

XX Claim 20; SEQ ID NO 26079; 1399pp + Sequence Listing: English.

PS The invention relates to human polynucleotides (AA179941-AA193841) and

XX the encoded proteins (AA000010-AA013910) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 60 AA:

Query Match 93.6%: Score 102; DB 22: Length 60;

Best Local Similarity 95.2%: Pred. No. 3.1e-05;

Matches 20: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CXXXXXXXXXXXXXXXXX 21

1 |XXXXXXXXXXXXXXXX|

DB 26 cXXXXXXXXXXXXXXXXX 46

Search completed: July 1, 2002, 06:19:22

Job time: 509 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 1, 2002, 06:28:08 ; Search time 49.45 Seconds

(without alignments)  
48,579 Million cell updates/sec

Title: US-09-461-684-2

Perfect score: 109  
Sequence: 1 CEAAAAAAAAAAAAAAAAAAAAA 25Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summariesDatabase : PIR-71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77	70.6	265	2	cgcr-4 protein - C
2	72	66.1	165	2	ribosomal protein
3	71	65.1	205	2	cgcr-1 protein - C
4	70	64.2	183	2	gene C98 protein - C
5	70	64.2	401	2	engrafted homeodom
6	69	63.3	97	2	antifreeze protein
7	69	63.3	1028	2	antifreeze protein
8	69	63.3	1213	2	DNA-binding protei
9	69	63.3	1668	2	ovo protein - frul
10	68	62.4	403	2	sex comb protein -
11	68	62.4	655	2	probable dihydrol
12	68	62.4	873	2	neurogenesis regul
13	68	62.4	2023	2	ecdysone-induced p
14	67	61.5	314	2	polycomb protein e
15	67	61.5	460	2	paired type homeob
16	67	61.5	494	2	hypothetical prote
17	67	61.5	497	2	zinc finger protei
18	67	61.5	604	2	myc-associated zin
19	67	61.5	606	2	homeotic protein B
20	67	61.5	627	2	Om(ID) protein - f
21	67	61.5	671	2	probable VME1 ATP-
22	67	61.5	1226	2	hypothetical prote
23	67	61.5	2125	2	eyeless - fruit fly
24	66.5	61.0	543	2	hypothetical prote
25	65.5	59.6	273	1	homeotic protein B
26	65	59.6	85	1	hypothetical prote
27	65	59.6	91	2	antifreeze protein
28	65	59.6	333	2	antifreeze protein
29	65	59.6	475	2	homeotic protein E

30	65	59.6	644	2	S39356	transcription fact
31	65	59.6	703	2	T48600	kinase-like protei
32	64	58.7	109	1	R6UP1	acidic ribosomal p
33	64	58.7	392	2	B48423	homeotic protein e
34	64	58.7	1065	2	T13230	dachshund isoform
35	64	58.7	1072	2	T13232	dachshund protein
36	64	58.7	1074	2	T13229	dachshund protein
37	64	58.7	1081	2	T13231	dachshund protein
38	64	58.7	1533	2	A46221	abdominal segment
39	64	58.7	2038	2	A43742	female sterile hom
40	63	57.8	179	2	F97683	50S ribosomal prot
41	63	57.8	179	2	AF2908	50S ribosomal prot
42	63	57.8	513	2	A48233	polyomavirus enhan
43	63	57.8	568	2	T39675	asparaginyl-trna s
44	63	57.8	581	2	E75383	conserved hypothet
45	63	57.8	846	2	S52418	GTP-binding regula

## ALIGNMENTS

RESULT 1  
S19113  
cgcr-4 protein - Chlamydomonas reinhardtii (fragment)  
C:Species: Chlamydomonas reinhardtii  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jul-2000  
R:Accession: S19113; S14466  
R:Wakarchuk, W.W.; Mueller, F.W.; Beck, C.F.  
Plant Mol. Biol. 18, 143-146, 1992  
A:Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrange  
A:Reference number: S19113; M01D:92119224  
A:Accession: S19113  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-265 <WAK>  
C:Cross-references: EMBL:X17208; NID:g18136; PIDN:CAJ5080.1; PID:g18137  
A:Gene: cgcr-4

Query Match 70.6% Score 77; DB 2; Length 265;  
Best Local Similarity 82.6% Pred. No. 0.28;  
Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAAAAAAAAAAAAA 25  
DB 154 AAAAAAAAAAKARVAAEARRAA 176

RESULT 2  
B87702  
ribosomal protein S16 (imported) - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
R:Accession: B87702  
R:Nierman, W.C.; Feldbylum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,  
n, J.; Ermolenko, M.; White, O.; Salberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4116-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; M01D:21173698; PMID:11259647  
A:Accession: B87702  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-165 <STO>  
A:Cross-references: GB:AE005673; NID:g13425408; PIDN:AAK25614.1; GSPDB:GNO0148  
C:Gene: CC3652

Query Match 66.1% Score 72; DB 2; Length 165;  
Best Local Similarity 75.0% Pred. No. 0.62;  
Matches 18; Conservative 2; Mismatches 4; Indels 0; Gaps 0;



```
RESULT 7
A:Binding protein ovo - fruit fly (Drosophila melanogaster)
A:Accession: A56038
C:Species: Drosophila melanogaster
C:Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
R:Garfinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.
Mol. Cell. Biol. 14, 6809-6818, 1994
A:Title: Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster
A:Reference number: A56038; MUID:95021209
A:Accession: A56038
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1028 <GAR>
A:Cross-references: GB:U11383; NID:9520526; PIDN:AAB60216.1; PID:9520527
C:Genetics:
A:Gene: ovo
A:Cross-references: FlyBase:FBgn0003028

Query Match 63.3%; Score 69; DB 2; Length 1028;
Best Local Similarity 85.7%; Pred. No. 4.6;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 AAAAAEAAAEEAAAEEAAA 23
Db 497 AAAAAAAAAAAAAAAAAAAAA 517

RESULT 8
S16356
ovo protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Feb-1997
R:Mevel-Ninio, M.; Terracol, R.; Kafatos, F.C.
EMBO J. 10, 2259-2266, 1991
A:Title: The ovo gene of Drosophila encodes a zinc finger protein required for female ge
A:Reference number: S16356; MUID:91293102
A:Accession: S16356
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1213 <MEV>
A:Cross-references: EMBL:X59772
C:Genetics:
A:Gene: FlyBase:ovo
A:Cross-references: FlyBase:FBgn0003028
A:Introns: 931/3; 1152/3

Query Match 63.3%; Score 69; DB 2; Length 1213;
Best Local Similarity 85.7%; Pred. No. 5.2;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 AAAAAEAAAEEAAAEEAAA 23
Db 860 AAAAAAAAAAAAAAAAAAAAA 880

RESULT 9
T13748
sex comb protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13748
R:Sincinall, D.A.R.; Milne, T.A.; Hodgson, J.W.; Sheppard, J.; Salinas, C.A.; Kyba, M.; F
Development 125, 1207-1216, 1998
A:Title: The Additional sex combs gene of Drosophila encodes a chromatin protein that b
A:Reference number: Z17750; MUID:9816384
A:Accession: T13748
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
```

```
A:Residues: 1-1668 <SIN>
A:Cross-references: EMBL:AJ001164; NID:g3292938; PIDN:CAA04568.1; PID:g3292939
C:Genetics:
A:Gene: fly
A:Cross-references: FlyBase:FBgn0000142
C:Function:
A:Description: Involved in repression of homeotic loci

Query Match 63.3%; Score 69; DB 2; Length 1668;
Best Local Similarity 78.3%; Pred. No. 6.6;
Matches 18; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAAAEAAAEEAAAEEAAA 24
Db 128 KAAAAAAAAAAAAAAAAAAAA 150

RESULT 10
A81882
Probable dihydrofolipamide S-succinyltransferase (EC 2.3.1.61) E2 component NMA1150 [1
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: A81882
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
A:Reference number: A81775; MUID:20222556
A:Accession: A81882
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-403 <PAR>
A:Cross-references: GB:A1162755; GB:A1157959; NID:g7379742; PIDN:CAB84412.1; PID:g737
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: sucB; NMA1150
A:Superfamily: dihydrofolipamide acetyltransferase; lipoyl/biotin-binding homology
C:Keywords: acyltransferase; coenzyme A

Query Match 62.4%; Score 68; DB 2; Length 403;
Best Local Similarity 75.0%; Pred. No. 2.9;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 EAAAAEAAAEEAAAEEAAA 25
Db 85 EAPAAATAAAEAPAAAPAA 108

RESULT 11
A29945
neurogenesis regulatory protein - fruit fly (Drosophila melanogaster) (fragment)
N:Alternate names: single-minded gene protein
C:Species: Drosophila melanogaster
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 20-Mar-1998
C:Accession: A29945
R:Crews, S.T.; Thomas, J.B.; Goodman, C.S.
Cell 52, 143-151, 1988
A:Title: The Drosophila single-minded gene encodes a nuclear protein with sequence si
A:Reference number: A29945; MUID:88151023
A:Accession: A29945
A:Molecule type: mRNA
A:Residues: 1-655 <CRE>
A:Cross-references: GB:M19020; NID:g158464; PID:g158465
C:Genetics:
A:Gene: sim
A:Cross-references: FlyBase:FBgn0004666
C:Keywords: DNA binding; transcription regulation

Query Match 62.4%; Score 68; DB 2; Length 655;
Best Local Similarity 66.7%; Pred. No. 4.2;
Matches 16; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
```



OY 2 EAAAAAEAAAAAEAAAAA 25  
:|||||:|||||  
Db 366 QAAQAQAQAQAQAQAQA 389

RESULT 12  
B53225  
ecdysone-induced protein E74A - fruit fly (Drosophila virilis)  
C:Species: Drosophila virilis  
C:Date: 12-May-1994 #sequence\_revision 12-May-1994 #text\_change 21-Feb-1997  
C:Accession: B53225  
R:Jones, C.W.; Dalton, M.W.; Townley, L.H.  
Genetics 127, 535-543, 1991  
A:Title: Interspecific comparisons of the structure and regulation of the Drosophila ecd  
A:Reference number: A53225; MUID:91200627  
A:Accession: B53225  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-873 <JON>  
A:Cross-references: CB:X59493  
C:Genetics:  
A:Gene: FlyBase: Dvir/Eip74EF  
A:Cross-references: FlyBase: FBgn0013076  
C:Superfamily: ets DNA-binding domain homology  
F:779-859/Domain: ets DNA-binding domain homology <ETS>

Query Match 62.4%; Score 68; DB 2; Length 873;  
Best Local Similarity 73.9%; Pred. No. 5.1;  
Matches 17; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAAEAAAAA 25  
:|||||:|||||  
Db 501 ASAAAAAAATHSAAAAA 523

RESULT 13  
T13154  
polycomb protein enhancer - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: T13154  
R:Stankunas, K.; Berger, J.; Ruse, C.; Sinclair, D.A.; Randazzo, F.; Brock, H.W.  
Development 125, 4053-4066, 1998  
A:Title: The enhancer of polycomb gene of Drosophila encodes a chromatin protein conserv  
A:Reference number: 217611; MUID:98407961  
A:Accession: T13154  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2023 <STA>  
A:Cross-references: EMBL: AF079764; NID: g3757889; PID: g3757890; PIDN: AAC64271.1  
A:Experimental source: Imaginal disc  
C:Genetics:  
A:Gene: E(Pc)  
A:Cross-references: FlyBase: FBgn0000581  
A:Map position: 2

Query Match 62.4%; Score 68; DB 2; Length 2023;  
Best Local Similarity 60.0%; Pred. No. 9.5;  
Matches 18; Conservative 3; Mismatches 3; Indels 6; Gaps 1;

OY 1 CE-----AAAAAEAAAAAEAAAAA 24  
:|||||:|||||  
Db 810 CEDOPVASTSAAAAAARAAEAAAAEAA 839

RESULT 14  
JC5273  
paired type homeobox protein, NBP - human  
C:Species: Homo sapiens (man)  
C:Date: 16-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 24-Sep-1999

C:Accession: JC5273  
R:Yokoyama, M.; Nishi, Y.; Yoshii, J.; Okubo, K.; Matsubara, K.  
DNA Res. 3, 311-320, 1996  
A:Title: Identification and cloning of neuroblastoma-specific and nerve tissue-speci  
A:Reference number: JC5272; MUID:97191543  
A:Contents: neuroblastoma cell  
A:Accession: JC5273  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-314 <YOK>  
A:Cross-references: DBJ: D82344; NID: g1841337; PIDN: BAA11555.1; PID: d1012222; PID: g1  
C:Comment: This protein is a transcriptional repressor involved in regulating gene e  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:99-155/Domain: homeobox homology <HOX>

Query Match 61.5%; Score 67; DB 2; Length 314;  
Best Local Similarity 78.3%; Pred. No. 3.1;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAAEAAAAA 25  
:|||||:|||||  
Db 244 AAAAAAAGGLAA 266

RESULT 15  
T33110  
hypothetical protein C18H7.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
C:Accession: T33110  
R:Tin-Wollam, A.; Fronick, W.  
submitted to the EMBL Data Library, May 1998  
A:Description: The sequence of C. elegans cosmid C18H7.  
A:Reference number: Z21284  
A:Accession: T33110  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-460 <TIN>  
A:Cross-references: EMBL: AF067607; PIDN: AC17641.1; GSDB: GN00022; CESP: C18H7.3  
A:Experimental source: strain Bristol N2; clone C18H7  
C:Genetics:  
A:Gene: CESP: C18H7.3  
A:Map position: 4  
A:Introns: 84/1  
C:Superfamily: Phaseolus glycyne-rich cell wall protein 1.8

Query Match 61.5%; Score 67; DB 2; Length 460;  
Best Local Similarity 62.5%; Pred. No. 4;  
Matches 20; Conservative 0; Mismatches 4; Indels 8; Gaps 1;

OY 2 EAAAAEAAAAE-----AAAAEAAAAA 25  
:|||||:|||||  
Db 399 EAAAAEAAAEAGAGGAEPGAPEAAAAA 430

Search completed: July 1, 2002, 06:28:10  
Job time: 712 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:31:33 ; Search time 25.37 Seconds  
(Without alignments)  
38.155 Million cell updates/sec

Title: US-09-461-684-2

Perfect score: 109  
Sequence: 1 CEAAAAAAAAAAAAAAAAAAAAA 25

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75.5	69.3	518	1	TPM4_DROME
2	72	66.1	165	1	RS16_CAUCR
3	70	64.2	183	1	OLEC_BRANA
4	70	64.2	401	1	HME1_MOUSE
5	69	63.3	97	1	ANP_LINFE
6	69	63.3	1028	1	OVO_DROME
7	68	62.4	1673	1	STM_DROME
8	68	62.4	1073	1	HR38_DROME
9	67	61.5	314	1	PXMB_MOUSE
10	67	61.5	314	1	PXMB_MOUSE
11	67	61.5	477	1	MA2_HUMAN
12	67	61.5	606	1	HMLD_DROAN
13	66	60.6	386	1	HXAD_MOUSE
14	66	60.6	388	1	HXAD_MOUSE
15	65	59.6	85	1	ANP4_PSEAM
16	65	59.6	91	1	ANP4_PSEAM
17	65	59.6	276	1	SK21_HUMAN
18	65	59.6	475	1	EVX2_MOUSE
19	65	59.6	476	1	EVX2_HUMAN
20	65	59.6	644	1	BTD_DROME
21	64	58.7	109	1	RLAI_TRYCR
22	64	58.7	392	1	HME1_HUMAN
23	64	58.7	1533	1	PUM_DROME
24	64	58.7	2038	1	FSH_DROME
25	63	57.8	376	1	FXE1_HUMAN
26	63	57.8	521	1	RUN2_HUMAN
27	63	57.8	590	1	HMAA_DROME
28	63	57.8	607	1	RUN2_MOUSE
29	63	57.9	1095	1	PIPA_DROME
30	62	56.9	364	1	HK61_MESAU
31	62	56.9	365	1	HK61_RAT
32	62	56.9	1355	1	SALM_DROME
33	61	56.0	31	1	ANP3_PAGBO

34	61	56.0	91	1	ANP4_PSEAM	P23699 pseudopleur
35	61	56.0	308	1	AEF1_DROME	P39413 drosophila
36	61	56.0	376	1	FXL2_HUMAN	P58012 homo sapien
37	61	56.0	421	1	BR3A_MOUSE	P17208 mus musculu
38	61	56.0	423	1	BR3A_HUMAN	001851 homo sapien
39	61	56.0	448	1	SPG1_STRSP	P06634 streptococc
40	61	56.0	593	1	SPG2_STRSP	P19909 streptococc
41	61	56.0	1733	1	VNUA_PRVKA	P33485 pseudorabie
42	60	55.0	74	1	SRI4_MACRA	018881 macaca radi
43	60	55.0	82	1	ANP4_PSEAM	P04002 pseudopleur
44	60	55.0	280	1	MACS_CHICK	P16527 gallus galli
45	60	55.0	367	1	HK61_HUMAN	P78426 homo sapien

## ALIGNMENTS

```

RESULT 1
ID TPM4_DROME STANDARD; PRT; 518 AA.
AC P49455;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Tropomyosin 1, fusion protein 33.
GN TMI OR TMII.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89127197; PubMed=2851721;
RA Hanke P.D., Storti R.V.;
RT "The Drosophila melanogaster tropomyosin II gene produces multiple
RT proteins by use of alternative tissue-specific promoters and
RT alternative splicing."
RL Mol. Cell. Biol. 8:3591-3602(1988).
RN [2]
RP SEQUENCE OF 1-286 FROM N.A.
RX MEDLINE=87064486; PubMed=3097506;
RA Karlik C.C., Fyrberg E.A.;
RT "Two Drosophila melanogaster tropomyosin genes: structural and
RT functional aspects."
RL Mol. Cell. Biol. 6:1965-1973(1986).
CC -1- ALTERNATIVE PRODUCTS: DROSOPHILA TROPOMYOSIN 1 GENE CAN PRODUCE
CC FOUR DIFFERENT ISOFORMS BY ALTERNATIVE PRODUCTS: A MUSCLE FORM,
CC A NON-MUSCLE FORM, AND TWO FUSION PROTEINS (33 AND 34).
CC -1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC -1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
CC -1- CAUTION: REF 2 SEQUENCE DIFFERS FROM THAT SHOWN MAINLY BECAUSE
CC OF ALTERNATIVE EXON USAGE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to a.license@sib-sib.ch).
CC -----
EMBL: X76208; CAA53800.1; -
EMBL: L00362; AAA28965.1; -
EMBL: M12840; AAA28965.1; JOINED.
EMBL: L00355; AAA28965.1; JOINED.
EMBL: L00356; AAA28965.1; JOINED.
EMBL: L00357; AAA28965.1; JOINED.
EMBL: L00358; AAA28965.1; JOINED.
EMBL: L00359; AAA28965.1; JOINED.
HSSP: P04002; IATP.
Flybase: FBgn0003721; Tm1.

```

DR InterPro: IPR000533; Tropomyosin.  
 DR Pfam: PF00261; Tropomyosin; 1.  
 DR PRINTS: PRO0194; TROPOMYOSIN.  
 DR PROSITE: PS00326; TROPOMYOSIN; 1.  
 KM Coiled coil; Repeat; Alternative splicing; Multigene family.  
 FT DOMAIN 14 267 COILED COIL (POTENTIAL).  
 FT DOMAIN 287 518 ALA/PRO-RICH.  
 FT CONFLICT 106 114 LGSATKLS -> SASAIQLAA (IN REF. 2).  
 FT CONFLICT 119 119 A -> S (IN REF. 2).  
 FT CONFLICT 183 183 A -> AMVEADLEAREEA (IN REF. 2).  
 FT CONFLICT 199 199 V -> L (IN REF. 2).  
 FT CONFLICT 215 231 NQREERKNQIKTLNTR -> TOKETTFETQIVLDHS (IN REF. 2).  
 SQ SEQUENCE 518 AA; 54558 MW; 153D0872CF9DB6EA CRC64;

Query Match 69.3%; Score 75.5; DB 1; Length 518;  
 Best Local Similarity 87.5%; Pred. No. 0.2;  
 Matches 21; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
 QY 2 EAAAAEAAAAEAAAAEAAAAA 25  
 Db 314 EAAAAEAAAAEAAAAEAAAAA 336

RESULT 2  
 RS16\_CAUCR STANDARD; PRT; 165 AA.  
 AC P58122;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 30S ribosomal protein S16.  
 GN RPS16 OR CC3652.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter.  
 NCBI\_TaxID=69394;  
 RN 1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nielsen W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eichen J., Heidelberg J.F., Alley M.R.K., Ohta N., Madcock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Ueberlack T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 CC -1- SIMILARITY: BELONGS TO THE S16P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE006023; AKK25614.1; -.  
 DR TIGR: CC3652;  
 DR InterPro: IPR000307; Ribosomal\_S16.  
 DR Pfam: PF00886; Ribosomal\_S16; 1.  
 DR PROSITE: PS00732; RIBOSOMAL\_S16; 1.  
 KM Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 165 AA; 17605 MW; ED46FC2798C5BE1C CRC64;

Query Match 66.1%; Score 72; DB 1; Length 165;  
 Best Local Similarity 75.0%; Pred. No. 0.18;  
 Matches 18; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAAAAEAAAAEAAAAEAAAAA 25  
 Db 115 QAEADAKAAAEAKAAAEAAAAA 138

RESULT 3  
 OLEC\_BRANA STANDARD; PRT; 183 AA.  
 AC P29526;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Oleosin C98 (Fragment).  
 GN C98.  
 OS Brassica napus (Rape).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Brassica.  
 NCBI\_TaxID=3708;  
 RN 1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Anther;  
 RX MEDLINE=93386188; PubMed=8374615;  
 RA Roberts M.R., Hodge R., Ross J.H.E., Sorensen A., Murphy D.J.,  
 RA Draper J., Scott R.;  
 RT "Characterization of a new class of oleosins suggests a male  
 RT gametophyte-specific lipid storage pathway.";  
 RL Plant J. 3:629-636(1993).  
 CC -1- FUNCTION: MAY HAVE A STRUCTURAL ROLE TO STABILIZE THE LIPID BODY  
 CC DURING DESICCATION OF THE SEED BY PREVENTING COALESCENCE OF THE  
 CC OIL. PROBABLY INTERACTS WITH BOTH LIPID AND PHOSPHOLIPID MOETIES  
 CC OF LIPID BODIES. MAY ALSO PROVIDE RECOGNITION SIGNALS FOR SPECIFIC  
 CC LIPASE ANCHORAGE IN LIPOLYSIS DURING SEEDLING GROWTH.  
 CC -1- SUBCELLULAR LOCATION: SURFACE OF OIL BODIES. OLEOSINS EXIST AT A  
 CC MONOLAYER LIPID/WATER INTERFACE.  
 CC -1- TISSUE SPECIFICITY: SPECIFIC TO THE MALE GAMETOPHYTE.  
 CC -1- SIMILARITY: BELONGS TO THE OLEOSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X67142; CAA47623.1; -.  
 DR PIR: S24960; S24960.  
 DR HSP: P04002; 1WFA.  
 DR InterPro: IPR000136; Oleosin.  
 DR Pfam: PF01277; Oleosin; 1.  
 DR PROSITE: PS00811; OLEOSINS; 1.  
 KM Seed, Oil body; Multigene family.  
 FT NON\_TER 1 1 POLAR.  
 FT DOMAIN <1 23 HYDROPHOBIC.  
 FT DOMAIN 24 95  
 SQ SEQUENCE 183 AA; 18149 MW; 198A5DB6DF3045A CRC64;

Query Match 64.2%; Score 70; DB 1; Length 183;  
 Best Local Similarity 78.3%; Pred. No. 0.31;  
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AAAAAEAAAAEAAAAEAAAAA 25  
 Db 151 AAPAAEPAPAAEAPAAEAPAA 173  
 RESULT 4  
 HME1\_MOUSE STANDARD; PRT; 401 AA.  
 ID HME1\_MOUSE  
 AC P09065;

01-NOV-1988 (Rel. 09, Created)  
01-FEB-1994 (Rel. 28, Last sequence update)  
16-OCT-2001 (Rel. 40, Last annotation update)  
Homeobox protein engrailed-1 (Mo-En-1).  
EN1 OR EN-1.  
Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE=93185339; PubMed=1363401;  
Logan C., Hanks M.C., Noble-Topham S., Nallathathan D.,  
Provart N.J., Joyner A.L.;  
"Cloning and sequence comparison of the mouse, human, and chicken  
engrailed genes reveal potential functional domains and regulatory  
regions.";  
Dev. Genet. 13:345-358(1992).  
[2]  
SEQUENCE OF 278-401 FROM N.A.  
RX MEDLINE=88112776; PubMed=2892757;  
Joyner A.L., Martin G.R.;  
"En-1 and En-2, two mouse genes with sequence homology to the  
Drosophila engrailed gene: expression during embryogenesis.";  
Genes Dev. 1:29-38(1987).  
[3]  
SEQUENCE OF 298-401 FROM N.A.  
RX MEDLINE=86079501; PubMed=2416459;  
Joyner A.L., Kornberg T., Coleman K.G., Cox D.R., Martin G.R.;  
"Expression during embryogenesis of a mouse gene with sequence  
homology to the Drosophila engrailed gene.";  
Cell 43:29-37(1985).  
[4]  
SEQUENCE OF 321-380 FROM N.A.  
RX MEDLINE=91099509; PubMed=1980115;  
Holland P.W.H., Williams N.A.;  
"Conservation of engrailed-like homeobox sequences during vertebrate  
evolution.";  
FEBS Lett. 277:250-252(1990).  
-1- SUBCELLULAR LOCATION: Nuclear.  
-1- SIMILARITY: BELONGS TO THE ENGRAILED FAMILY OF HOMEBOX PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL: L12703; AAA03660.2; -  
DR EMBL: Y00201; CAA68361.1; -  
DR PIR: A26629; A26629.  
DR PIR: A24778; A24778.  
DR PIR: S13009; S13009.  
DR PIR: A48423; A48423.  
DR HSSP: P02636; 3HDD.  
DR TRANSFAC: T02016; -  
DR MGD: MGT:95389; En1.  
DR InterPro: IPR000747; Engrailed.  
DR InterPro: IPR001356; Homeobox.  
DR Pfam: PF00046; homeobox.1.  
DR PRINTS: PR00026; ENGRAILED.  
DR PRINTS: PR00024; HOMEBOX.  
DR SMART: SM00389; HOX.1.  
DR PROSITE: PS00027; HOMEBOX\_1; 1.  
DR PROSITE: PS50071; HOMEBOX\_2; 1.  
DR PROSITE: PS00033; ENGRAILED; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.  
FT DOMAIN 52 87 PRO-RICH.  
FT DOMAIN 73 87 POLY-PRO.  
FT DOMAIN 207 228 POLY-ALA.  
FT DNA\_BIND 312 371 HOMEBOX.

SQ SEQUENCE 401 AA; 40950 MW; 1F90210950152FAE CRC64;  
Query Match 64.2%; Score 70; DB 1; Length 401;  
Best Local Similarity 78.3%; Pred. No. 0.57;  
Matches 18; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 3 AAAAAEAAAAEAAAAEAAAAA 25  
DB 207 AAAAAAAAAAAAAAAAAAAS 229  
RESULT 5  
ANP LIMFE STANDARD; PRT; 97 AA.  
ID ANP LIMFE  
AC P09031;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 01-AUG-1990 (Rel. 15, Last annotation update)  
DE Antifreeze protein precursor (AFP).  
OS Limanda ferruginea (Yellowtail flounder).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;  
OC Pleuronectoidae; Pleuronectidae; Limanda.  
NCBI\_TaxID=8258;  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE=88029483; PubMed=3665937;  
Scott G.K., Davies P.L., Shears M.A., Fletcher G.L.;  
"Structural variations in the alanine-rich antifreeze proteins of the  
pleuronectinae.";  
Eur. J. Biochem. 168:629-633(1987).  
-1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.  
-1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE I AFP ARE  
ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL: X06356; CAA29655.1; -  
DR PIR: S02376; S02376.  
DR HSSP: P04002; IAFP.  
DR InterPro: IPR000104; Antifreeze\_1.  
DR PRINTS: PR00308; ANTIFREEZE1.  
KW Antifreeze protein; Repeat; Signal.  
FT SIGNAL 1 23  
FT PROPEP 24 48  
FT CHAIN 49 97  
FT SEQUENCE 97 AA; 8865 MW; 62AD582DF8E459B6 CRC64;  
Query Match 63.3%; Score 69; DB 1; Length 97;  
Best Local Similarity 70.8%; Pred. No. 0.24;  
Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 2 EAAAAEAAAAEAAAAEAAAAA 25  
DB 53 DAAAAAATATAAAKAAADTAAAA 76  
RESULT 6  
OVO\_DROME STANDARD; PRT; 1028 AA.  
AC P51521; O9XZU4;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, last annotation update)  
DE OVO protein (Shaven baby protein).  
OS OVO OR SVB.  
OC Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=ovary;  
RX MEDLINE=95021209; PubMed=7935398;  
RA Garlinkel M.D., Wang J., Liang Y., Mahowald A.P.;  
RT "Multiple products from the shavenbaby-ovo gene region of Drosophila  
melanogaster: relationship to genetic complexity.";  
RL Mol. Cell. Biol. 14:6809-6818(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OREGON-R;  
RX MEDLINE=91293102; PubMed=1712294;  
RA Mevel-Ninio M.T.M., Teriacol R., Kafatos F.C.;  
RT "The ovo gene of Drosophila encodes a zinc finger protein required  
for female germ line development.";  
RL EMBO J. 10:2259-2266(1991).  
CC -1- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM  
CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -1- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND  
ACCUMULATES IN NURSE CELLS DURING OOCYTESIS. STORED IN THE EGG,  
BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED  
PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U11383; AAB60216.1; -  
DR EMBL: X59772; CAB36921.1; ALT\_SEQ.  
DR HSSP: p25490; 12NM.  
DR Flybase: Fgn0003028; ovo.  
DR InterPro: IPR000822; Znf-C2H2.  
DR Pfam: PF00096; zf-C2H2; 4.  
DR PRINTS: PRO0048; ZINC\_FINGER.  
DR SMART: SM00355; ZNF\_C2H2; 4.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 3.  
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;  
KW Transcription regulation.  
KW  
FT DOMAIN 62 66 POLY-ALA.  
FT DOMAIN 72 77 POLY-GLY.  
FT DOMAIN 80 85 POLY-GLY.  
FT DOMAIN 98 108 POLY-GLY.  
FT DOMAIN 144 152 POLY-HIS.  
FT DOMAIN 153 159 POLY-ASN.  
FT DOMAIN 336 339 POLY-GLN.  
FT DOMAIN 347 353 POLY-GLN.  
FT DOMAIN 357 361 POLY-GLN.  
FT DOMAIN 410 414 POLY-GLN.  
FT DOMAIN 418 422 POLY-GLN.  
FT DOMAIN 426 432 POLY-GLN.  
FT DOMAIN 445 453 POLY-GLN.  
FT DOMAIN 456 459 POLY-GLN.  
FT DOMAIN 466 474 POLY-ALA.  
FT DOMAIN 517 524 POLY-SER.  
FT DOMAIN 529 549 POLY-ALA.  
FT DOMAIN 558 581 POLY-ALA.  
FT DOMAIN 639 651 POLY-ALA.  
FT DOMAIN 717 725 POLY-ALA.  
FT DOMAIN 797 802 POLY-GLN.

FT DOMAIN 820 823 POLY-GLN.  
FT DOMAIN 826 832 POLY-GLN.  
FT DOMAIN 874 992 ZINC\_FINGERS.  
FT ZN\_FING 874 896 C2H2-TYPE.  
FT ZN\_FING 902 924 C2H2-TYPE.  
FT ZN\_FING 930 953 C2H2-TYPE.  
FT ZN\_FING 969 992 C2H2-TYPE.  
FT CONFLICT 647 647 A -> R (IN REF. 2).  
SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2BC0F6F77 CRC64;  
  
Query Match 63.3%; Score 69; DB 1; Length 1028;  
Best Local Similarity 85.7%; Pred. No. 1.5;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Caps 0;  
  
QY 3 AAAAAEAAAAEAAAAEAAA 23 -  
DB 497 AAAAAAAAAAAAAAAAAAAAAA 517  
  
RESULT 7  
SIM\_DROME STANDARD; PRT; 673 AA.  
AC P05709; O96521; Q9VEZ3;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Single-minded protein.  
GN SIM OR CG7771.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE OF 19-673 FROM N.A.  
RX MEDLINE=88151023; PubMed=3345560;  
RA Crews S.T., Thomas J.B., Goodman C.S.;  
RT "The Drosophila single-minded gene encodes a nuclear protein with  
sequence similarity to the per gene product.";  
RL Cell 52:143-151(1988).  
RN [2]  
RP SEQUENCE OF 1-18 FROM N.A., AND SIMILARITY TO HUH PROTEINS.  
RX MEDLINE=92103681; PubMed=1760843;  
RA Nambu J.R., Lewis J.O., Wharton K.A. Jr., Crews S.T.;  
RT "The Drosophila single-minded gene encodes a helix-loop-helix protein  
that acts as a master regulator of CNS midline development.";  
RL Cell 67:1157-1167(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99054545; PubMed=9840810;  
RA Kasai Y., Stahl S., Crews S.;  
RT "Specification of the Drosophila CNS midline cell lineage: direct  
control of single-minded transcription by dorsal/ventral patterning  
genes.";  
RL gene Expr. 7:171-189(1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Ceolniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,  
RA Adair J.F., Agbayani A., An H.-J., Andrews-Planck C., Baldwin D.,  
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,



RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Domnes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Eosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Godec A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iegem C.,  
 RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Morkulov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*";  
 RA Science 287:2185-2195(2000).  
 RL [5]  
 RN SEQUENCE OF 528-1073 FROM N.A.  
 RP STRAIN-CANTON-S;  
 RC MEDLINE-96068664; PubMed-7479849;  
 RX Fluk G.U., Thummel C.S.;  
 RA "Isolation, regulation, and DNA-binding properties of three  
 RT Drosophila nuclear hormone receptor superfamily members.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:10604-10608(1995).  
 CC -1- FUNCTION: BINDS TO NGFI-B RESPONSE ELEMENTS. PLAYS AN IMPORTANT  
 CC ROLE IN LATE STAGES OF EPIDERMAL METAMORPHOSIS.  
 CC -1- SUBUNIT: FORMS A HETERODIMER WITH USP.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED IN PREBLASTODERM  
 CC EMBRYOS, SPECIFICALLY IN CENTRAL NERVOUS SYSTEM AND INTESTINAL  
 CC TRACT. HIGHLY EXPRESSED IN THIRD INSTAR LARVAL IMAGINAL DISKS AND  
 CC BRAIN COMPLEXES, BUT NOT IN OVARIES.  
 CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN 0-8 HOUR EMBRYOS AND ADULTS.  
 CC HIGHER IN LATE EMBRYOGENESIS AND DURING LARVAL AND PUPAL STAGES.  
 CC SHORT ISOFORM IS ENRICHED IN PUPAE AND ADULTS, LONG ISOFORM IN  
 CC LARVAE.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NR4 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X89246; CAA61534.1; -  
 CC EMBL: Y15606; CAA75690.1; -  
 CC EMBL: AJ002073; CAA05172.1; -  
 CC EMBL: AE003667; AAF53914.1; -  
 CC EMBL: U36762; AAC64926.1; -  
 CC HSSP: P19793; 2NLL.  
 CC TRANSFAC: T02760; -  
 CC FLYBASE: Fggn0014859; Ht38.  
 CC InterPro: IPR000536; Hormone\_rec\_119.  
 CC InterPro: IPR001723; Stridhormone\_receptor.  
 CC InterPro: IPR001628; zf-C4.  
 CC Pfam: PF00104; hormone\_rec; 2.  
 DR Pfam: PF00105; zf-C4; 2.

DR PRINTS: PR00398; STRDHORMONER.  
 DR PRINTS: PR00047; STROIDFINGER.  
 DR SMART: SM00430; HNF\_C4; 1.  
 DR SMART: SM00399; znf-C4; 1.  
 DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; 1.  
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 FT Zinc-finger; Alternative splicing; Developmental protein.  
 FT DNA\_BIND 744 809  
 FT 2N\_BIND 744 764  
 FT 2N\_FING 780 804  
 FT 2N\_FING 188 804  
 FT DOMAIN 188 192  
 FT DOMAIN 206 218  
 FT DOMAIN 221 228  
 FT DOMAIN 268 272  
 FT DOMAIN 294 312  
 FT DOMAIN 441 462  
 FT DOMAIN 505 508  
 FT DOMAIN 619 626  
 FT DOMAIN 661 665  
 FT VARSPIC 1 522  
 FT CONFLICT 667 667  
 FT CONFLICT 685 685  
 FT CONFLICT 689 692  
 FT CONFLICT 697 697  
 FT CONFLICT 702 702  
 FT CONFLICT 1041 1041  
 FT CONFLICT 1064 1064  
 SEQUENCE 1073 AA; 116991 MW; 126A30DAFA1C096A CRC64;  
 Query Match 62.4%; Score 68; DB 1; Length 1073;  
 Best Local Similarity 73.9%; Pred. No. 1.9;  
 Matches 17; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Oy 3 AAAAEEAAAAEAAAAEAAAAA 25  
 Db 206 AATTAATAAAEAGGAASAAAA 228  
 RESULT 9  
 PMXB\_HUMAN  
 ID PMXB\_HUMAN STANDARD; PRT; 314 AA.  
 AC Q99453;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Paired mesoderm homeobox protein 2b (Paired-like homeobox 2b)  
 DE (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBphox).  
 GN PMXB2B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Neuroblastoma;  
 RX MEDLINE-97191543; PubMed-9039501;  
 RA Yokoyama M., Nishi Y., Yoshii J., Okubo K., Matsubara K.;  
 RT "Identification and cloning of neuroblastoma-specific and nerve  
 RT tissue-specific genes through compiled expression profiles";  
 RL DNA Res. 3:311-320(1996).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE-20487360; PubMed-11034547;  
 RA Adachi M., Browne D., Lewis E.J.;  
 RT "Paired-like homeodomain proteins Phox2a/Arx and Phox2b/NBphox have  
 RT beta-hydroxylase gene transcription";  
 RL DNA Cell Biol. 19:539-554(2000).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE-99326521; PubMed-10395798;  
 RA Yokoyama M., Matanabe H., Nakamura M.;



```

RT "Genomic structure and functional characterization of NBphox (PMX2B),
RT a homeodomain protein specific to catecholaminergic cells that is
RT involved in second messenger-mediated transcriptional activation.";
RL Genomics 59:40-50(1999).
CC
CC -1- FUNCTION: INVOLVED IN THE DEVELOPMENT OF SEVERAL MAJOR
CC NORADRENERGIC NEURON POPULATIONS, INCLUDING THE LOCUS COERULEUS.
CC TRANSCRIPTION FACTOR WHICH COULD DETERMINE A NEUROTRANSMITTER
CC PHENOTYPE IN VERTEBRATES. ENHANCES SECOND-MESSENGER-MEDIATED
CC ACTIVATION OF THE DOPAMINE BETA-HYDROLYASE AND C-FOS PROMOTERS,
CC AND OF SEVERAL ENHANCERS INCLUDING CYCLIC AMP-RESPONSE ELEMENT AND
CC SERUM-RESPONSE ELEMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEUROBLASTOMA, BRAIN AND ADRENAL
CC GLAND.
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
DR EMBL: D82344; BAA1555.1; -
DR EMBL: AF117979; AAD2698.1; -
DR EMBL: AB015671; BAA82670.1; -
DR HSSP: P14653; 1B72.
DR MIM: 603851; -
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; Homeobox.1.
DR PRINTS: PR00024; HOMEBOX.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
KM Homeobox: DNA-binding: Developmental protein: Nuclear protein:
KW Transcription regulation.
FT DNA_BIND 98 157 HOMEBOX.
FT DOMAIN 159 167 POLY-ALA.
FT DOMAIN 212 217 POLY-GLY.
FT DOMAIN 241 260 POLY-ALA.
SO SEQUENCE 314 AA; 31607 MW; 76737F71948B5D81 CRC64;

Query Match 61.5%; Score 67; DB 1; Length 314;
Best Local Similarity 78.3%; Pred. No. 0.95;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAAEEAAAAA 25
DB 244 AAAAAAAAAAAAAAGGLAAA 266

RESULT 10
PMXB_MOUSE STANDARD: PRT; 314 AA.
AC 035690;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)
DE (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).
GN PMX2B OR PHOX2B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98040559; PubMed=9374403;
RA Patlyn A., Morin X., Cremer H., Goridis C., Brunet J.-F.;
RT "Expression and interactions of the two closely related homeobox
RT genes Phox2a and Phox2b during neurogenesis.";

```

```

RL Development 124:4065-4075(1997).
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=99326521; PubMed=10395798;
RA Yokoyama M., Matanabe H., Nakamura M.;
RT "Genomic structure and functional characterization of NBphox (PMX2B),
RT a homeodomain protein specific to catecholaminergic cells that is
RT involved in second messenger-mediated transcriptional activation.";
RL Genomics 59:40-50(1999).
CC
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
DR EMBL: Y14493; CAAT4833.1; -
DR EMBL: AB015672; BAA82671.1; -
DR HSSP: P06601; 1FJL.
DR MGD: MGI:1100882; Pmx2b.
DR InterPro: IPR000047; HTH_repressr.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; Homeobox.1.
DR PRINTS: PR00024; HOMEBOX.
DR PRINTS: PR00031; HTHREPRESSR.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
KM Homeobox: DNA-binding: Developmental protein: Nuclear protein.
KW Homeobox.
FT DNA_BIND 98 157 HOMEBOX.
FT DOMAIN 159 167 POLY-ALA.
FT DOMAIN 212 217 POLY-GLY.
FT DOMAIN 241 260 POLY-ALA.
SO SEQUENCE 314 AA; 31621 MW; 40737F71948B595A CRC64;

Query Match 61.5%; Score 67; DB 1; Length 314;
Best Local Similarity 78.3%; Pred. No. 0.95;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAAEEAAAAA 25
DB 244 AAAAAAAAAAAAAAGGLAAA 266

RESULT 11
MAZ_HUMAN STANDARD: PRT; 477 AA.
AC P56270; Q99443; Q15703;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myc-associated zinc finger protein (MAZ1) (Purine-binding
DE transcription factor) (Pur-1) (ZF87) (ZIF87).
GN MAZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92366479; PubMed=1502157;
RA Bossone S.A., Asselin C., Patel A.J., Marcu K.B.;
RT "MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences
RT regulating transcriptional initiation and termination.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7452-7456(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Carcinoma;

```

RX MEDLINE-92232709; PubMed-1567856;  
 RA Pyrc J.J., Moberg K.H., Hall D.J.;  
 RT "Isolation of a novel cDNA encoding a zinc-finger protein that binds  
 to two sites within the c-myc promoter.";  
 RL Biochemistry 31:4102-4110(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Pancreatic islets;  
 RX MEDLINE-96428591; PubMed-8831693;  
 RA Tsutsui H., Sakatsume O., Itakura K., Yokoyama K.K.;  
 RT "Members of the MAZ family: a novel cDNA clone for MAZ from human  
 pancreatic islet cells.";  
 RL Biochem. Biophys. Res. Commun. 226:801-809(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-96224025; PubMed-8626793;  
 RA Parks C.L., Shenk T.;  
 RT "The serotonin 1a receptor gene contains a TATA-less promoter that  
 responds to MAZ and Sp1.";  
 RL J. Biol. Chem. 271:4417-4430(1996).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Lymphoblastoma;  
 RX MEDLINE-98352105; PubMed-9685418;  
 RA Song J., Murakami H., Tsutsui H., Tang X., Matsumura M., Itakura K.,  
 Kanazawa I., Sun K., Yokoyama K.K.;  
 RT "Genomic organization and expression of a human gene for Myc-  
 associated zinc finger protein (MAZ).";  
 RL J. Biol. Chem. 273:20603-20614(1998).  
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR WITH DUAL ROLES  
 IN TRANSCRIPTION INITIATION AND TERMINATION. BINDS TO TWO SITES,  
 MEI1 AND MEI2, WITHIN THE C-MYC PROMOTER HAVING GREATER  
 AFFINITY FOR THE FORMER. ALSO BINDS TO MULTIPLE G/C-RICH SITES  
 WITHIN THE PROMOTER OF THE SPI FAMILY OF TRANSCRIPTION FACTORS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL  
 MUSCLE, AND PANCREAS. SEEMS NOT TO BE EXPRESSED IN KIDNEY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M94046; -; NOT ANNOTATED. CDS.  
 DR EMBL: M93339; -; NOT ANNOTATED. CDS.  
 DR EMBL: D85131; BA12728.1; ALT\_INIT.  
 DR EMBL: U33819; AAB04121.1; ALT\_INIT.  
 DR EMBL: AB017335; BAA33064.1; -;  
 DR HSSP: P08046; 1AAY.  
 DR TRANSFAC: T00490; -;  
 DR TRANSFAC: T02305; -;  
 DR MIM: 600999; -;  
 DR InterPro: IPR000822; znf-C2H2.  
 DR Pfam: PF00096; zf-C2H2; 6.  
 DR PRINTS: PR00048; ZINCFIN2.  
 DR SMART: SM00355; znf\_C2H2; 6.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_2; 5.  
 DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 5.  
 KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;  
 KW RNA-binding; Repeat; Nuclear protein.  
 FT DOMAIN 190 413 ZINC\_FINGERS.  
 FT ZN\_FING 190 212 C2H2-TYPE.  
 FT ZN\_FING 279 301 C2H2-TYPE.  
 FT ZN\_FING 307 329 C2H2-TYPE.  
 FT ZN\_FING 337 360 C2H2-TYPE.  
 FT ZN\_FING 366 388 C2H2-TYPE.  
 FT ZN\_FING 392 413 C2H2-TYPE.  
 FT DOMAIN 96 108 POLY-ALA.  
 FT DOMAIN 133 139 POLY-PRO.  
 FT DOMAIN 157 161 POLY-ALA.

FT DOMAIN 245 249 POLY-GLY.  
 FT DOMAIN 435 449 MISSING (IN REF. 3).  
 FT CONFLICT 259 259 L -> M (IN REF. 2 AND 4).  
 FT CONFLICT 401 401 MISSING (IN REF. 3).  
 FT CONFLICT 443 447 MISSING (IN REF. 3).  
 SQ SEQUENCE 477 AA; 48607 MW; C04C80F32C3C6825 CRC64;

Query Match 61.58; Score 67; DB 1; Length 477;  
 Best Local Similarity 77.38; Pred. No. 1.3;  
 Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 AAAAEEAAAEEAAAEEAAA 25  
 DB 90 AAAGSAAAAAAAVAA 111

RESULT 12  
 HMID\_DROAN STANDARD; PRT; 606 AA.

AC P22544;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Homeobox protein OM(ID).  
 GN OM(ID).  
 OS Drosophila ananassae (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7217;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91122048; PubMed-1671353;  
 RA Tanda S., Corces V.G.;  
 RT "Retroltransposon-Induced overexpression of a homeobox gene causes  
 defects in eye morphogenesis in Drosophila.";  
 RL EMBO J. 10:407-417(1991).  
 CC -1- FUNCTION: Probably involved in eye morphogenesis.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X56682; CAA40011.1; -;  
 DR PIR: S13367; S13367.  
 DR HSSP: P22808; 1VND.  
 DR FlyBase: FBgn0012114; DanaB.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox; 1.  
 DR PRINTS: PR00024; HOMEBOX.  
 DR SMART: SM00389; HOX; 1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 KW DNA-binding; Homeobox; Developmental protein; Nuclear protein; Vision.

KW DNA-binding; Homeobox; Developmental protein; Nuclear protein; Vision.  
 FT DOMAIN 23 57 HIS/GLN-RICH (OPA-REPEAT).  
 FT DOMAIN 106 124 HIS/PRO-RICH (OPA-REPEAT).  
 FT DOMAIN 173 193 HIS/PRO-RICH.  
 FT DNA\_BIND 331 390 HOMEBOX.  
 FT DOMAIN 220 248 ALA-RICH.  
 FT DOMAIN 422 434 ALA-RICH.  
 FT DOMAIN 455 455 ALA-RICH.  
 FT DOMAIN 503 510 ALA-RICH.  
 FT DOMAIN 515 521 PRO-RICH.  
 SQ SEQUENCE 606 AA; 61735 MW; AA7BB8B367370FBB CRC64;

Query Match 61.5%; Score 67; DB 1; Length 606;  
 Best Local Similarity 78.3%; Pred. No. 1.6;  
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAAAAA 25  
 ||||| ||| ||||| |||||  
 Db 220 AAAAAAAAAAAAAAAAAAAAAA 242

RESULT 13  
 HXAD\_MOUSE ID STANDARD: PRT: 386 AA.  
 AC 062424;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Homeobox protein Hox-A13 (Hox-1.10).  
 GN HOXA13 OR HOX-1.10.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96259555; PubMed=8673126;  
 RA Mortlock D.P., Post L.C., Innis J.W.;  
 RT "The molecular basis of hypodactyly (Hd): a deletion in Hoxa 13 leads  
 to arrest of digital arch formation.";  
 RL Nat. Genet. 13:284-289(1996).  
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
 A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
 SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS (BY  
 SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DISEASE: DEFECTS IN HOXA13 ARE THE CAUSE OF HYPODACTYLY (HD), A  
 CONDITION CHARACTERIZED BY PROFOUND DEFICIENCY OF DIGITAL ARCH  
 STRUCTURES.  
 CC -1- SIMILARITY: BELONGS TO THE ABD-B FAMILY OF HOMEOBOX PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U59322; AAB03322.1; -;  
 CC HSSP: P02833; ISAN.  
 CC TRANSFAC: T03337; -;  
 DR MGD: MGI:96173; Hoxa13.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox; 1.  
 DR SMART: SM00389; HOX; 1.  
 DR PROSITE: PS00027; HOMEOBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEOBOX\_2; 1.  
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation.  
 KM  
 FT DOMAIN 38 51 POLY-ALA.  
 FT DNA\_BIND 320 379 HOMEOBOX.  
 FT DOMAIN 52 57 POLY-GLY.  
 FT DOMAIN 62 66 POLY-ALA.  
 FT DOMAIN 73 84 POLY-ALA.  
 FT DOMAIN 101 104 POLY-ALA.  
 FT DOMAIN 116 133 POLY-ALA.  
 FT DOMAIN 198 205 POLY-ALA.  
 SQ SEQUENCE 386 AA; 39566 MW; 2B01DC9B1951324 CRC64;

Query Match 60.6%; Score 66; DB 1; Length 386;  
 Best Local Similarity 70.8%; Pred. No. 1.4;  
 Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25  
 || : ||||| ||||| |||||  
 Db 111 EAPPSAAAAAAAAAAAAAAAAAAS 134

RESULT 14  
 HXAD\_HUMAN ID STANDARD: PRT: 388 AA.  
 AC P31271; O43371;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Homeobox protein Hox-A13 (Hox-1J).  
 GN HOXA13 OR HOX1J.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97172976; PubMed=9020844;  
 RA Mortlock D.P., Innis J.W.;  
 RT "Mutation of HOXA13 in hand-foot-genital syndrome.";  
 RL Nat. Genet. 15:179-180(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Bradsaw H., Hinds K., Keppler D.;  
 RT Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 322-387 FROM N.A.  
 RX MEDLINE=90098876; PubMed=2574852;  
 RA Acampora D., D'Esposito M., Fajella A., Pannese M., Migliaccio E.,  
 RA Morelli F., Stornaiuolo A., Nigro V., Simone A., Boncinelli E.;  
 RT "The human HOX gene family".  
 RL Nucleic Acids Res. 17:10385-10402(1989).  
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
 A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
 SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: BELONGS TO THE ABD-B FAMILY OF HOMEOBOX PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U82827; AAC50993.1; -;  
 CC EMBL: AC004080; -; NOT\_ANNOTATED\_CDS.  
 CC PIR: S14932; S14932.  
 CC HSSP: P02833; ISAN.  
 CC TRANSFAC: T03321; -;  
 DR MIM: 142959; -;  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox; 1.  
 DR SMART: SM00389; HOX; 1.  
 DR PROSITE: PS00027; HOMEOBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEOBOX\_2; 1.  
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation.  
 KM  
 FT DNA\_BIND 322 381 HOMEOBOX.  
 FT DOMAIN 38 53 POLY-ALA.  
 FT DOMAIN 62 66 POLY-ALA.  
 FT DOMAIN 73 84 POLY-ALA.  
 FT DOMAIN 116 133 POLY-ALA.  
 FT DOMAIN 145 150 POLY-ALA.  
 FT DOMAIN 200 207 POLY-ALA.  
 FT DOMAIN 146 146 POLY-ALA.  
 FT CONFLICT 146 146 A -> G (IN REF. 2).  
 FT CONFLICT 187 187 P -> H (IN REF. 2).  
 FT CONFLICT 195 195 P -> A (IN REF. 2).  
 FT CONFLICT 198 198 P -> A (IN REF. 2).

SQ SEQUENCE 388 AA; 39752 MW; 6CD9C9A5616C2FF6 CRC64;

## Query Match

Best Local Similarity 60.6%; Score 66; DB 1; Length 388;  
Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25

DB 111 EAPPSAAAAAAAAAAAAAAAAAS 134

## RESULT 15

ANP4\_PSEAM

ID ANP4\_PSEAM STANDARD; PRT; 85 AA.

AC P02734;

DT 21-JUL-1986 (Rel. 01, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Antifreeze peptide 4 precursor.

OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes americanus).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes; Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.

OC NCBI\_TaxID=8265;

OX [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=81247379; PubMed=6265915;

RA Lin Y., Gross J.K.;

RT "Molecular cloning and characterization of winter flounder antifreeze

CDNA";

RL Proc. Natl. Acad. Sci. U.S.A. 78:2825-2829(1981).

CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.

CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE

CC ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: J00930; AAA49467.1; -.

DR PIR: A03193; FDPLAW.

DR HSSP: P04002; 1WFA.

DR InterPro: IPR000104; Antifreeze\_1.

DR PRINTS: PR00308; ANTIFREEZE1.

KW Antifreeze protein; Repeat; Multigene family; Signal.

FT SIGNAL 1 21

FT CHAIN 22 85

FT SIGNAL 1 21

FT CHAIN 22 85

FT SIGNAL 1 21

FT CHAIN 22 85

FT SIGNAL 1 21

FT CHAIN 22 85

FT SIGNAL 1 21

FT CHAIN 22 85

FT SIGNAL 1 21

FT CHAIN 22 85

Search completed: July 1, 2002, 06:31:34  
Job time: 671 sec



RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Moberly C., Morris J., Mostrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003698; AAF5488.2; -  
DR FlyBase: FBgn0038108; CG7518.  
DR InterPro: IPR001005; Myb\_DNA\_bind.  
DR PROSITE: PS00037; MYB\_1; UNKNOWN\_1.  
SQ SEQUENCE 2451 AA; 266959 MW; 088A2293F27481E2 CRC64;

Query Match 77.1%; Score 84; DB 5; Length 2451;  
Best Local Similarity 91.3%; Pred. No. 1.5;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAEAAAAEAAAAA 25  
DB 1374 AAAAAEAAAAEAAAAEAAAAA 1396

RESULT 2  
ID Q9EPW8 PRELIMINARY; PRT; 1354 AA.  
AC Q9EPW8;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE NISCHARIN.  
GN NISCH.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ICR OUTBRED; TISSUE=BRAIN;  
RX MEDLINE=20571837; PubMed=11121431;  
RA Alahari S.K., Lee J.W., Juliano R.L.;  
RT "Nischarin, a Novel Protein That Interacts with the Integrin  $\alpha$ has  
RT Subunit and Inhibits Cell Migration.";  
RL J. Cell Biol. 151:1141-1154(2000).  
DR EMBL: AF315344; AAG42100.1; -  
DR MGD: MGI:1928323; Nisch.  
DR InterPro: IPR001128; Cyt\_P450.  
DR InterPro: IPR001611; LRR.  
DR InterPro: IPR003592; LRR\_out.  
DR Pfam: PF00560; LRR\_4.  
DR PRINTS: PR00019; LEURICHRPT.  
DR SMART: SM00370; LRR\_5.  
DR PROSITE: PS00086; CYTOCHROME\_P450; UNKNOWN\_1.  
SQ SEQUENCE 1354 AA; 148060 MW; 01BD676FDCA19247 CRC64;

Query Match 73.4%; Score 80; DB 11; Length 1354;  
Best Local Similarity 83.3%; Pred. No. 2.2;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAAAEAAAAEAAAAEAAAAA 25  
DB 837 EAAAAEAAAAEAAAAEAAAAA 860

RESULT 3  
ID Q91MW0 PRELIMINARY; PRT; 110 AA.  
AC Q91MW0;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 10.7 KDA PROTEIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM "N.A."  
RC TISSUE=SALIVARY GLAND;  
RA Strausberg R.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC012681; AAH12681.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 110 AA; 10662 MW; 0581D2F635F87EAB CRC64;

Query Match 72.5%; Score 79; DB 11; Length 110;  
Best Local Similarity 83.3%; Pred. No. 0.32;  
Matches 20; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 EAAAAEAAAAEAAAAEAAAAA 25  
DB 23 DAAAAEAAAAEAAAAEAAAAA 46

RESULT 4  
ID Q39598 PRELIMINARY; PRT; 265 AA.  
AC Q39598;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE CGCR-4 PROUCT (FRAGMENT).  
GN CGCR-4.  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3053;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92119224; PubMed=1731966;  
RA Wakarchuk W.W., Muller F.W., Beck C.F.;  
RT "Two GC-rich DNA elements of *Chlamydomonas reinhardtii* with complex  
RT arrangements of directly repeated sequence motifs.";  
RL Plant Mol. Biol. 18:143-146(1992).  
DR EMBL: X17208; CAA35080.1; -  
FT NON\_TER 1 1  
SQ SEQUENCE 265 AA; 26216 MW; B35318B7377CF782 CRC64;

Query Match 70.6%; Score 77; DB 10; Length 265;  
Best Local Similarity 82.6%; Pred. No. 1;  
Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAEAAAAEAAAAA 25  
DB 154 AAAAAEAAAAEAAAAEAAAAA 176

RESULT 5  
ID Q90XG2 PRELIMINARY; PRT; 665 AA.  
AC Q90XG2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE CHOROIDEREMIA PROTEIN.  
 GN CHM.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C;  
 RA van den Hurk J.A., Huber I., van de Pol T.J., Cremers F.P.;  
 RT Cloning and sequencing of the mouse choroideremia gene."  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF218084; AAF25478.1; -.  
 DR MGD; MGI:892979; Chm.  
 DR InterPro: IPR002005; Rab\_GDI\_RBP.  
 DR Pfam; PF00996; GDI; 1.  
 DR PRINTS; PRO0891; RABGDI\_RP.  
 SO SEQUENCE 665 AA; 73976 MW; FF71A74AD3BDE0A CRC64;

Query Match 69.7%; Score 76; DB 11; Length 665;  
 Best Local Similarity 79.2%; Pred. No. 2.8;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAEAAAAAEAAAAAEAAAAA 25  
 DB 132 EAAEAAEAATEAAEAAEAAEAA 155

RESULT 6  
 ID 09H4A0 PRELIMINARY; PRT; 1452 AA.  
 AC 09H4A0;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CDC2L5 PROTEIN KINASE.  
 GN CDC2L5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PLACENTA;  
 RA Marques F., Moreau J.L., Peaucellier G., Lozano J.C., Schatt P.,  
 RA Picard A., Callebaut I., Perre E., Genevriere A.M.;  
 RT "A new subfamily of high molecular mass CDC2-related kinases with  
 RT PITAI/VRE."  
 RL Biochem. Biophys. Res. Commun. 279:832-837(2001).  
 DR EMBL; AJ297710; CAC10401.1; -.  
 DR HSSP; P24941; 1BUH.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR SMART; SM00219; TYKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW ATP-binding; Kinase; Transferase  
 SO SEQUENCE 1452 AA; 158480 MW; C7ED0729688439CB CRC64;

Query Match 69.7%; Score 76; DB 4; Length 1452;  
 Best Local Similarity 79.2%; Pred. No. 5.5;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAEAAAAAEAAAAAEAAAAA 25  
 DB 463 EAAEAAEAATAEATKAEEAAAKA 486

RESULT 7  
 ID 09H4A1 PRELIMINARY; PRT; 1512 AA.  
 AC 09H4A1;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CDC2L5 PROTEIN KINASE.  
 GN CDC2L5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PLACENTA;  
 RA Marques F., Moreau J.L., Peaucellier G., Lozano J.C., Schatt P.,  
 RA Picard A., Callebaut I., Perre E., Genevriere A.M.;  
 RT "A new subfamily of high molecular mass CDC2-related kinases with  
 RT PITAI/VRE."  
 RL Biochem. Biophys. Res. Commun. 279:832-837(2001).  
 DR EMBL; AJ297709; CAC10400.1; -.  
 DR HSSP; P24941; 1BUH.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR SMART; SM00219; TYKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW ATP-binding; Kinase; Transferase.  
 SO SEQUENCE 1512 AA; 164969 MW; 283B8D553DB57650 CRC64;

Query Match 69.7%; Score 76; DB 4; Length 1512;  
 Best Local Similarity 79.2%; Pred. No. 5.7;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAEAAAAAEAAAAAEAAAAA 25  
 DB 463 EAAEAAEAATAEATKAEEAAAKA 486

RESULT 8  
 ID 024426 PRELIMINARY; PRT; 531 AA.  
 AC 024426;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE TROPOMYOSIN ISOFORM 33.  
 GN TMI OR CG4898.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-84205681; PubMed-6202423;  
 RA Karlik C.C., Mahaffey J.W., Coutu M.D., Fyrberg E.A.;  
 RT "Organization of contractile protein genes within the 88F subdivision  
 RT of the D. melanogaster third chromosome."  
 RL Cell 37:469-481(1984).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-87064486; PubMed-3097506;  
 RA Karlik C.C., Fyrberg E.A.;  
 RT "Two Drosophila melanogaster tropomyosin genes: structural and  
 RT functional aspects."  
 RL Mol. Cell. Biol. 6:1965-1973(1986).

DR EMBL: K02621; AAA28968.1; JOINED.  
DR EMBL: M12840; AAA28968.1; JOINED.  
DR EMBL: L00355; AAA28968.1; JOINED.  
DR EMBL: L00356; AAA28968.1; JOINED.  
DR EMBL: L00357; AAA28968.1; JOINED.  
DR EMBL: L00358; AAA28968.1; JOINED.  
DR EMBL: L00359; AAA28968.1; JOINED.  
DR EMBL: L00360; AAA28968.1; JOINED.  
DR EMBL: L00362; AAA28968.1; JOINED.  
DR FlyBase: FBgn0003721; Tm1.  
DR InterPro: IPR002965; P-rich\_extensn.  
DR Pfam: PF00261; Tropomyosin; 2.  
DR PRINTS: PR01217; PRICHEXTENS.  
DR PROSITE: PS00326; TROPOMYOSIN.  
KW Alternative splicing; TROPOMYOSIN; 1.  
SQ SEQUENCE 531 AA; 56047 MW; 2618A715E20EADFC CRC64;

Query Match: 69.3%; Score 75.5; DB 5; Length 531;  
Best Local Similarity 87.5%; Pred. No. 2.6;  
Matches 21; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 2 EAAAAAFAAAAAEAAAAEAAAA 25  
DB 327 EAAAAAFAAAAAEAAAAEAAAA 349

RESULT 9  
O9VF97 PRELIMINARY; PRT; 566 AA.  
AC O9VF97;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Tm1 PROTEIN.  
GN Tm1 OR CG4898.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abail J.F., Agbayani A., An H.-J., Andrews-Planckoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Morklov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclieb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zavert J.S., Zhan M., Zhang G., Zhao O., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003708; AAF5164.1; Tm1.  
DR FlyBase: FBgn0003721; Tm1.  
DR InterPro: IPR002965; P-rich\_extensn.  
DR InterPro: IPR000533; Tropomyosin.  
DR Pfam: PF00261; Tropomyosin; 2.  
DR PRINTS: PR01217; PRICHEXTENS.  
DR PROSITE: PS00326; TROPOMYOSIN.  
SQ SEQUENCE 566 AA; 60449 MW; 2AF07CF80DA2032E CRC64;

Query Match: 69.3%; Score 75.5; DB 5; Length 566;  
Best Local Similarity 87.5%; Pred. No. 2.8;  
Matches 21; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 2 EAAAAAFAAAAAEAAAAEAAAA 25  
DB 362 EAAAAAFAAAAAEAAAAEAAAA 384

RESULT 10  
O9VXD3 PRELIMINARY; PRT; 221 AA.  
AC O9VXD3;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE CG13012 PROTEIN.  
GN CG13012.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abail J.F., Agbayani A., An H.-J., Andrews-Planckoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,



RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003503; AAF4641.1; -;  
 DR Flybase: FBgn0030769; CG13012.  
 SQ SEQUENCE 221 AA; 22987 MW; A1B95919B167C5E2 CRC64;

Query Match 68.8%; Score 75; DB 5; Length 221;  
 Best Local Similarity 81.8%; Pred. No. 1.4;  
 Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 AAAAAAAAAAAAAAAAAAAAAA 25  
 DB 6 AAAAAAAAAAATVAATAA 27

RESULT 11  
 ID 09W2J2 PRELIMINARY; PRT; 1071 AA.  
 AC 09W2J2:  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE CG18375 PROTEIN.  
 GN CG18375.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RC MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dubin K.J., Evangelista C.C., Ferraz C., Ferrelira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Clodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Jimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu Q., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern R., Pacle J.M.,

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003453; AAF4699.1; -;  
 DR HSSP: Q13625; IYCS.  
 DR Flybase: FBgn0034606; CG18375.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00023; ank; 2.  
 DR Pfam: PF00018; SH3; 1.  
 DR SMART: SM00248; ANK; 2.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS50088; ANK\_REPEAT; 2.  
 DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 KW ANK repeat; Repeat.  
 SQ SEQUENCE 1071 AA; 115502 MW; F86840E92692B4FA CRC64;

Query Match 66.1%; Score 72; DB 5; Length 1071;  
 Best Local Similarity 79.2%; Pred. No. 10;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25  
 DB 504 EAAAAAAAAAAAAAAAAAAAAA 527

RESULT 12  
 ID 039597 PRELIMINARY; PRT; 206 AA.  
 AC 039597;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE CGCR-1 PRODUCT (FRAGMENT).  
 GN CGCR-1.  
 OS *Chlamydomonas reinhardtii*.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadales; Chlamydomonas.  
 RX NCBI\_TaxID=3055;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CW15 MT-;  
 RC MEDLINE=92119224; PubMed=1731966;  
 RA Wakarchuk W.W., Muller F.W., Beck C.F.;  
 RT "Two GC-rich DNA elements of *Chlamydomonas reinhardtii* with complex  
 RL arrangements of directly repeated sequence motifs.";  
 DR Plant Mol. Biol. 18:143-146(1992).  
 DR EMBL: X17207; CAA35079.1; -;  
 DR InterPro: IPR001778; POA\_allergen.  
 DR PRINTS: PR00833; POAALLERGM.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 206 AA; 19869 MW; ED3FF120EF8EFAE1 CRC64;

Query Match 65.1%; Score 71; DB 10; Length 206;  
 Best Local Similarity 75.0%; Pred. No. 3.1;  
 Matches 18; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25  
 DB 49 EAAAAAAAAAAAAAAAAAAAAA 72

	RESULT	13	
ID	O9SMH3	PRELIMINARY;	PRT; 1216 AA.
AC	O9SMH3:		
DT	01-MAY-2000 (TREMBlrel_13, Created)		
DR	01-MAY-2000 (TREMBlrel_13, last sequence update)		
DT	01-DEC-2001 (TREMBlrel_19, last annotation update)		
DE	VARIABLE FLAGELLAR NUMBER PROTEIN.		
CN	VFLI.		
OS	Chlamydomonas reinhardtii.		
OC	Eukaryota: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;		
CC	Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.		
OX	NCBITaxID=3055;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-21GR, CC-1690;		
RX	MEDLINE=94063478; PubMed=8244002;		
RA	Tam L.W., Lefebvre P.A.;		
RT	"Cloning of flagellin genes in Chlamydomonas reinhardtii by DNA		
RL	insertional mutagenesis.";		
RM	Genetics 135::375-384(1993).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-21GR, CC-1690;		
RX	MEDLINE=21181843; PubMed=11285274;		
RA	Silflow C.D., LaVoie M., Tam L.W., Tousey S., Sanders M., Wu W.C.,		
Borodovsky M., Lefebvre P.A.;			
RT	"The Vll protein in Chlamydomonas localizes in a rotationally		
RL	asymmetric pattern at the distal ends of the basal bodies.";		
J. Cell Biol.	153:63-74(2001).		
DR	EMBL; AF154916; AAD52203.1.-.		
IPIPR004089; Chemotaxis_transducer.			
InterPro; IPR001611; LRR.			
DRI	InterPro; IPR003592; LRR_out.		
PFfam; PF00560; LRR; 5.			
SMART; SM00370; LRR; 3.			
KW	Flagella.		
SO	SEQUENCE	1216 AA; 127943 MW; F06E798B35AF256E CRC64;	
	Query Match	65.1%; Score 71; DB 10; Length 1216;	
	Best Local Similarity	73.9%; Pred. NO.14;	
	Matches 17; Conservative 1; Mismatches 5; Indels 0; Gaps 0.		
QY	2 EAAATAAEEAAAEEAAAEEAAAA 24		
DB	635 EAAGREAAAAGVQAAMAEELAVA 657		
RESULT	14		
O9M4X9			
ID	O9M4X9	PRELIMINARY;	PRT; 1787 AA.
AC	O9M4X9:		
DT	01-OCT-2000 (TREMBLrel_15, Created)		
DR	01-OCT-2000 (TREMBLrel_15, last sequence update)		
DT	01-DEC-2001 (TREMBLrel_19, last annotation update)		
DE	FLAGELLAR AUTOTOMY PROTEIN PAIP.		
CN	PAI.		
OS	Chlamydomonas reinhardtii.		
OC	Eukaryota: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;		
CC	Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.		
OX	NCBITaxID=3055;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	Finst R.J., Kim P.J., Griffiths E.R., Quarby L.M.;		
RT	"Paip is a 171 kDa protein essential for axonomal microtubule severing		
RL	in Chlamydomonas.";		
J. Cell Sci.	0:0-0(2000).		
DR	EMBL; AF246590; AAT66419.1.-.		
IPIPR003592; LRR_out.			
SMART; SM00370; LRR; 7.			

sq	SEQUENCE	1787	AA;	171584	MM;	452A74259EE14DC1	CRC64;
	Query Match		65.1%;	Score 71;	DB 10;	Length 1787;	
	Best Local Similarity		81.8%;	Pred. No. 20;			
	Matches 18;	Conservative	0;	Mismatches	4;	Indels	0;
0y	3	AAAAAAAAAAAAAAAAAAAA	24				
Db	1704	AAAAAAAAAAAAAAAAAAAA	1725				

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
5	AAAEAAAAAAAEAAAAA	25				
141	AAAEAAAAVVAAAAAAA	161				

64.2% Score 70; DB 4; Length 323;  
85.7% Pred. No. 5.6;  
0; Mismatches 3; Indels 0; Gaps 0;

RESULT 15

09H782

PRELIMINARY: PRT: 323 AA.

ID 09H782

AC 09H782

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE CDNA: FLJ21157 FIS, CLONE CAS09937 (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxId=9606;

RL [1]

RP SEQUENCE FROM N.A.

RA Watanabe K., Kunagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,

RA Suzuki Y., Ohashi M., Nishi T., Shibahara T., Tanaka T.,

RA Nakamura Y., Isegai T., Sugano S.;

RT "NEDO human cDNA sequencing project."

RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL; AK024810; BAB15016.1; -

DR InterPro: IPR000104; Antifreeze\_1.

DR InterPro: IPR002713; FF.

DR InterPro: IPR001202; WW.

DR Pfam: PF01846; FF; 1.

DR PRINTS: PRO0308; ANTIFREEZE1.

DR SMART: SM00441; FF; 1.

DR SMART: SM00456; WW; 2.

DR PROSITE: PS01159; WW\_DOMAIN\_1; 1.

DR PROSITE: PS50020; WW\_DOMAIN\_2; 2.

DR NON TER 323

FT 323

SEQUENCE 323 AA: 35148 MW; EAC7B36489FA074A CR:64;

Search completed: July 1, 2002, 06:30:57  
Job time: 689 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:22 ; Search time 98.97 Seconds  
(Without alignments)  
28.057 Million cell updates/sec

Title: US-09-461-684-2

Perfect score: 109  
Sequence: 1 CEAAAAAAAAAAAAAAAAAAAAA 25

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

A.Geneseq-032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109	100.0	25	21	AA13781
2	109	100.0	45	21	AA13783
3	84	77.1	2451	22	ABB71574
4	79	72.5	119	22	ABB40362
5	79	72.5	119	22	AA661172
6	79	72.5	119	22	AA673886
7	76	69.7	112	22	AA634068
8	76	69.7	112	22	ABG15501
9	75	69.3	566	22	AA605268
10	75	69.3	566	22	ABB61040
11	75	68.8	221	22	ABB67690

12	73	67.0	225	21	AA600758	Human secreted pro
13	72	66.1	54	18	AA622875	P. americanus skin
14	72	66.1	316	22	ABG18917	Novel human diagno
15	72	66.1	1071	22	ABB60579	Drosophila melanog
16	72	66.1	1444	22	ABG15667	Novel human diagno
17	70	64.2	39	18	AA622874	P. americanus skin
18	70	64.2	326	22	AA660765	Gene 16 related pe
19	70	64.2	326	22	AA660766	Gene 16 related pe
20	70	64.2	487	22	AA641402	Human polypeptide
21	70	64.2	957	22	AA64150	Human WPI-Interact
22	69	63.3	39	18	AA622876	Synthetic skin typ
23	69	63.3	39	18	AA622881	P. americanus anti
24	69	63.3	262	22	ABG15586	Novel human diagno
25	69	63.3	1529	17	AA67985	CORC potassium cha
26	69	63.3	1669	22	ABG64003	Drosophila melanog
27	68	62.4	39	18	AA622871	P. americanus skin
28	68	62.4	39	18	AA622869	P. americanus skin
29	68	62.4	148	22	AB670649	Drosophila melanog
30	68	62.4	1078	22	AB658620	Drosophila melanog
31	68	62.4	2023	22	AB653457	Drosophila melanog
32	67	61.5	314	18	AA614283	Human neuroblastom
33	67	61.5	314	21	AA670572	Mouse Plox2b prote
34	67	61.5	544	22	AB651602	Drosophila melanog
35	67	61.5	1037	22	AB670288	Drosophila melanog
36	67	61.5	1416	22	AB654499	Drosophila melanog
37	67	61.5	2703	22	AB653299	Drosophila melanog
38	66	60.6	76	17	AA608166	Peptide modulating
39	66	60.6	31	21	AA691273	Elmeria gametocyte
40	66	60.6	76	20	AA627190	Amino acid fragmen
41	66	60.6	92	21	AA644712	Shorthorn sculpin
42	66	60.6	372	22	AB64062	Human protein sequ
43	66	60.6	339	22	AB63773	Human protein sequ
44	66	60.6	634	22	AB671624	Drosophila melanog
45	66	60.6	924	22	AB667870	Drosophila melanog

## ALIGNMENTS

RESULT 1	
AA613781	
ID	AA613781 standard; peptide: 25 AA.
XX	
AC	AA613781:
XX	
DT	10-NOV-2000 (first entry)
XX	
DE	Soluble peptide antigen pEA.
XX	
KW	pEA peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW	major histocompatibility complex class I; MHC class I; antigen; tumour;
KW	prostate; breast; multiple myeloma.
XX	
OS	Unidentified.
XX	
PN	WO200035949-A1.
XX	
PD	22-JUN-2000.
XX	
PF	14-DEC-1999: 99MO-US29724.
XX	
PR	14-DEC-1998: 98US-0112324.
XX	
PA	(DEND-) DENDREON CORP.
XX	
PI	Laus R, Hakim I, Vidovic D;
XX	
DR	WPI: 2000-442365/38.
XX	
PT	Antigens modified by the covalent addition of a peptide that
PT	facilitates entry into antigen presenting cells, useful for producing
XX	compositions for immunizing against tumors and pathogens -



```
RESULT 4
ABBA40362
ID ABBA40362 standard; Peptide: 119 AA.
XX
XX
AC ABBA40362;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #7868 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-483447/52.
XX
DR WPI: 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 27: SEQ ID NO 32997; 639pp + sequence listing; English.
XX
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 119 AA;

Query Match 72.5%; Score 79; DB 22; Length 119;
Best Local Similarity 83.3%; Pred. No. 0.0043;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAAAAEEAAAAAEAAAAA 25
DB 31 eaeaeaaaaaeaaaaaeaaaa 54

RESULT 5
AAM61172
ID AAM61172 standard; Protein: 119 AA.
XX
XX
AC AAM61172;
XX
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 33277.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX
```

```
KW epilepsy; cancer.
XX
XX OS Homo sapiens.
XX
XX PN WO200157275-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00667.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-483446/52.
XX
DR WPI: 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX PS Example 4: SEQ ID NO: 33277; 650pp + Sequence listing; English.
XX
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
XX
SQ Sequence 119 AA;

Query Match 72.5%; Score 79; DB 22; Length 119;
Best Local Similarity 83.3%; Pred. No. 0.0043;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAAAAEEAAAAAEAAAAA 25
DB 31 eaeaeaaaaaeaaaaaeaaaa 54

RESULT 6
AAM73886
ID AAM73886 standard; Protein: 119 AA.
XX
XX
AC AAM73886;
XX
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 34192.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00668.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
```

PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488900/53.  
XX

PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -

XX  
PS Example 4; SEQ ID NO: 34192; 658bp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention.  
XX

SO Sequence 119 AA;

Query Match 72.5%; Score 79; DB 22; Length 119;  
Best Local Similarity 83.3%; Pred. No. 0.0043;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAAAAEAAAAAEAAAAAEAAAA 25

Db 31 eaeaeaeaeaeaeaeaeaeaeaeae 54

RESULT 7

ID AAM34068 standard; Protein: 119 AA.

XX AAM34068;

DT 17-OCT-2001 (first entry)

DE Peptide #8105 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

KM genetic disorder.

OS Homo sapiens.

XX WO200157272-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488897/53.  
DR  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human placenta -  
XX

PS Claim 27; SEQ ID No 34337; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;  
CC see AAI31315-AA157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders.  
XX

SO Sequence 119 AA;

Query Match 72.5%; Score 79; DB 22; Length 119;  
Best Local Similarity 83.3%; Pred. No. 0.0043;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAAAAEAAAAAEAAAAAEAAAA 25

Db 31 eaeaeaeaeaeaeaeaeaeaeaeae 54

RESULT 8

ID ABG15501 standard; Protein: 112 AA.

XX ABG15501;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #15492.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.  
N-PSDB; AAS79688.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX

PS Claim 20; SEQ ID No 45860; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 112 AA:

Query Match 69.7%; Score 76; DB 22; Length 112;

Best Local Similarity 87.0%; Pred. No. 0.0095;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAEAAAAEAAAAA 25

||||| ||||| ||||| |||||

Db 82 aaaaaaaaaaavaaaaaaaaaa 104

RESULT 9

AAR05268  
ID AAR05268 standard; protein; 737 AA.

XX  
AC AAR05268;

DT 15-AUG-1990 (first entry)

DE Amino acid sequence of human megakaryocytopoietin (MKP)

DE encoded by upper reading frame of DNA contained in clone 14.

XX Human megakaryocytopoietin (MPK); haematopoiesis; immunoassay;

KM rare leukaemia diagnosis.

XX Homo sapiens.

OS EP354989-A.

PN JP03195496-A.

XX 21-FEB-1990.

PD 27-JUN-1989; 89JP-0111714.

XX 28-JUN-1988; 88US-0212623.

XX (YISS ) YISSUM RES DEV CO.

XX Hermona S, Haim Z;

PI WPI; 1990-052749/08.

DR WPI; 1991-292630/40.

DR N-PSDB; AAQ92263.

XX New polypeptide hormone megakaryocytopoietin -

PT involved in haematopoiesis and immune response, and derived

PT nucleic acid, antibodies, etc., useful eg in diagnosis of leukaemia

XX Claim 7; Page 17; Fig 1; 24pp; English.

XX A cDNA library constructed from foetal ganglioside mRNA was screened  
CC with a butyryl-cholinesterase (bCh) cDNA probe to identify one clone  
CC (clone 14) contg. about 250 bases at the 5'-region of the bCh coding  
CC sequence plus a sequence which hybridised with genomic DNA from patients  
CC with acute myelocytic leukaemia. AARP5268 is the translation of the  
CC upper reading frame and is claimed in the patent. Its nucleic acid can  
CC be isolated, opt. together with its signal sequence, by screening cDNA  
CC or genomic libraries with the clone 14 DNA, to identify a full-length  
CC clone. The material from this clone can be transferred into mammalian or  
CC microbial host cells and these cultured for its prodn. It has hormonal  
CC activity in modulation of haematopoiesis and immune responses. It is  
CC useful in standard immunoassay or hybridisation procedures for  
CC classification and diagnosis of rare leukaemias. It may also be useful  
CC therapeutically.

XX  
SQ Sequence 737 AA:

Query Match 69.7%; Score 76; DB 11; Length 737;

Best Local Similarity 79.2%; Pred. No. 0.061;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAEAAAAEAAAAEAAAAA 25

||||| ||||| ||||| |||||

Db 106 eaataaatakaatakaataaka 129

RESULT 10

ABB61040  
ID ABB61040 standard; protein; 566 AA.

XX  
AC ABB61040;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 9912.

XX Drosophila: developmental biology; cell signalling; insecticide;

KM pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

DR N-PSDB; ABL05143.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Disclosure; SEQ ID NO 9912; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 566 AA:

Query Match 69.3%; Score 75.5; DB 22; Length 566;

Best Local Similarity 87.3%; Pred. No. 0.054;

Matches 21; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 2 EAAAAEAAAAEAAAAEAAAAA 25

||||| ||||| ||||| |||||

Db 362 eaataaataaataa-aaataaataa 384





```
XX (HSCR-) HSC RES & DEV LP.
XX
XX Gong 2, Hew C;
XX
XX WPI: 1997-402614/37.
XX DR N-PSDB: AAT75505.
XX
XX Isolated fish skin anti-freeze polypeptide - useful to depress
XX freezing points of aqueous compositions and protect plant, fungal,
XX animal or bacterial cells from cold
XX
XX Claim 6: Page 68; 104pp; English.
XX
XX This sequence represents the sAFP8 clone corresponding to a novel skin
XX intracellular antifreeze polypeptide (AFP) which induces a concentration
XX dependent decrease in the freezing point of an aqueous solution. This
XX novel polypeptide does not contain a signal sequence and is thought to be
XX intracellular. AFP's can be used to make an aqueous composition resistant
XX to freezing by changing its thermal hysteresis such as a water or salt
XX solution, an intracellular compartment of a cell or a food stuff, e.g.
XX soft serve "frozen" yoghurt or ice cream. AFP's can inhibit ice
XX recrystallisation during cold storage, improving the texture and
XX palatability of the food and has antibacterial properties. Such
XX polypeptides can also be expressed to provide cold resistance to cells,
XX e.g. plant, fungal animal or bacterial cells. The antibodies can be used
XX to identify and isolate AFP while its promoter can be used to direct
XX expression of a nucleic acid.
XX
XX Sequence 54 AA:
SQ
Query Match 66.1%; Score 72; DB 18; Length 54;
Best Local Similarity 78.3%; Pred. No. 0.014;
Matches 18; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 3 AAAAAEAAAAAEAAAAEAAAAA 25
DB 28 aaaaatacaaaakaaaakaaaaa 50
RESULT 14
ABG18917
ID ABG18917 standard; Protein: 316 AA.
XX
XX ABG18917;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #18908.
XX
XX Human: Chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX DR N-PSDB: AAS83104.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
```

```
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 49276; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 316 AA:
SQ
Query Match 66.1%; Score 72; DB 22; Length 316;
Best Local Similarity 82.6%; Pred. No. 0.083;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 3 AAAAAEAAAAAEAAAAEAAAAA 25
DB 14 aaaaaaavaaaaaaaavaaaaaa 36
RESULT 15
ABB60579
ID ABB60579 standard; Protein: 1071 AA.
XX
XX ABB60579;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 8529.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
XX DR N-PSDB: ABL04682.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
```

PT Interactions -  
 XX  
 PS Disclosure: SEQ ID NO 8529; 21pp + Sequence Listing: English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
 CC sequences (ABB57737-ABB72072).  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1071 AA:

Query Match 66.1%; Score 72; DB 22; Length 1071;  
 Best Local Similarity 79.2%; Pred. No. 0.28;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 EAAAAAEEAAAAEEAAAA 25  
 ||||| ||||| ||||| |||||  
 Db 504 eaaaaaaaaaaaaaaaaaaaaa 527

Search completed: July 1, 2002, 06:19:23  
 Job time: 510 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 1, 2002, 06:28:10 ; Search time 49.45 Seconds

(without alignments)  
46.636 Million cell updates/sec

Title: US-09-461-684-3

Sequence: 143  
1 CGLFGAIGFIENGWEGMIDGMYG 24Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

PIR-71:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	93.7	550	1 HMIYS2	hemagglutinin prec
2	134	93.7	550	1 HMIYS3	hemagglutinin prec
3	134	93.7	550	1 HMIY77	hemagglutinin prec
4	134	93.7	550	1 HMIY80	hemagglutinin prec
5	134	93.7	550	1 HMIY33	hemagglutinin prec
6	134	93.7	550	1 HMIY89	hemagglutinin prec
7	134	93.7	550	1 HMIY21	hemagglutinin prec
8	134	93.7	550	1 HMIY15	hemagglutinin prec
9	134	93.7	550	1 HMIY98	hemagglutinin prec
10	134	93.7	550	2 J01153	hemagglutinin prec
11	134	93.7	550	2 J01154	hemagglutinin prec
12	134	93.7	550	2 J01155	hemagglutinin prec
13	134	93.7	556	1 HMIYVH	hemagglutinin prec
14	134	93.7	556	1 HMIYVH	hemagglutinin prec
15	134	93.7	556	1 HMIYVH	hemagglutinin prec
16	134	93.7	556	1 HMIYDU	hemagglutinin prec
17	133	93.0	561	1 HMIY49	hemagglutinin prec
18	133	93.0	561	1 HMIY84	hemagglutinin prec
19	132	92.3	565	1 HMIY61	hemagglutinin prec
20	132	92.3	565	1 HMIY61	hemagglutinin prec
21	132	92.3	566	1 HMIY61	hemagglutinin prec
22	132	92.3	566	1 HMIY61	hemagglutinin prec
23	131	91.6	550	2 S38637	hemagglutinin prec
24	131	91.6	550	1 HMIY86	hemagglutinin prec
25	131	91.6	560	1 HMIYV7	hemagglutinin prec
26	131	91.6	565	1 HMIY62	hemagglutinin prec
27	131	91.6	565	1 HMIY64	hemagglutinin prec
28	131	91.6	565	1 HMIY65	hemagglutinin prec
29	131	91.6	565	1 HMIY66	hemagglutinin prec

30	131	91.6	565	1 HMIY67	hemagglutinin prec
31	131	91.6	565	1 HMIY68	hemagglutinin prec
32	131	91.6	565	1 HMIY69	hemagglutinin prec
33	131	91.6	565	1 HMIY69	hemagglutinin prec
34	131	91.6	565	1 HMIY69	hemagglutinin prec
35	131	91.6	565	2 S33703	hemagglutinin prec
36	131	91.6	570	1 A45591	hemagglutinin - in
37	131	91.6	570	2 S22013	hemagglutinin prec
38	131	91.6	570	2 S22013	hemagglutinin prec
39	131	91.6	570	2 S22015	hemagglutinin prec
40	131	91.6	570	2 S22016	hemagglutinin prec
41	131	91.6	570	2 S22017	hemagglutinin prec
42	131	91.6	570	2 S22018	hemagglutinin prec
43	131	91.6	570	2 S22020	hemagglutinin prec
44	131	91.6	570	2 S22021	hemagglutinin prec
45	131	91.6	570	2 S22029	hemagglutinin prec

## ALIGNMENTS

RESULT 1  
HMIYS2  
hemagglutinin precursor - Influenza A virus (strain A/swine/126/82) (fragment)  
C:Species: Influenza A virus  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 18-Sep-1998  
C:Accession: A29971  
R:Kida, H.; Shortridge, K.F.; Webster, R.G.  
Virology 162, 160-166, 1988  
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in Chir  
A:Reference number: A94370; M0ID:88101364  
A:Accession: A29971  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M19056; NID:g324208  
A>Note: the sequence in Genbank entry FLAHAP, release 106, (PID:g324209) differs frc  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: Influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7% Score 134; DB 1; Length 550;

Best local similarity 100.0% Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 2 CGLFGAIGFIENGWEGMIDGMYG 24  
DB 330 CGLFGAIGFIENGWEGMIDGMYG 352

## RESULT 2

HMIYS3  
hemagglutinin precursor - Influenza A virus (strain A/swine/81/78) (fragment)  
C:Species: Influenza A virus  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 18-Sep-1998  
C:Accession: B29971  
R:Kida, H.; Shortridge, K.F.; Webster, R.G.  
Virology 162, 160-166, 1988  
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in Chln  
A:Reference number: A94370; M0ID:88101364  
A:Accession: B29971  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M19057; NID:g324210  
A>Note: the sequence in Genbank entry FLAHAPB, release 106, (PID:g324211) differs frc  
C:Genetics:

A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domains: transmembrane #status predicted <TM1>  
F:8-22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIEGMEGMDGMYG 24  
|||||  
Db 330 GLFGAIAGFIEGMEGMDGMYG 352

RESULT 3  
HMI177  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: A27813  
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: A27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16737; NID:g324081; PIDN:AAA43143.1; PID:g324082  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domains: transmembrane #status predicted <TM1>  
F:8-22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIEGMEGMDGMYG 24  
|||||  
Db 330 GLFGAIAGFIEGMEGMDGMYG 352

RESULT 4  
HMI180  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: B27813  
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: B27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16738; NID:g324083  
A:Note: the translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1  
C:Genetics:

A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domains: transmembrane #status predicted <TM1>  
F:8-22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIEGMEGMDGMYG 24  
|||||  
Db 330 GLFGAIAGFIEGMEGMDGMYG 352

RESULT 5  
HMI173  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: C27813  
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: C27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16739; NID:g324085; PIDN:AAA43145.1; PID:g324086  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domains: transmembrane #status predicted <TM1>  
F:8-22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIEGMEGMDGMYG 24  
|||||  
Db 330 GLFGAIAGFIEGMEGMDGMYG 352

RESULT 6  
HMI189  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: D27813  
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: D27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16740; NID:g324087; PIDN:AAA43146.1; PID:g324088  
C:Genetics:  
A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Product: hemagglutinin HA2 #status predicted <TM1>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8.22.38.165.285.483/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAFGIENGWEGMIDGMYG 24  
Db 330 GLFGAIAFGIENGWEGMIDGMYG 352

RESULT 7  
HMIY21  
hemagglutinin precursor - Influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: E27813  
R:Kida, H.; Kawaoaka, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: E27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16741; NID:g324089  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: Influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Product: hemagglutinin HA2 #status predicted <TM1>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8.22.38.165.285.483/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAFGIENGWEGMIDGMYG 24  
Db 330 GLFGAIAFGIENGWEGMIDGMYG 352

RESULT 8  
HMIY98  
hemagglutinin precursor - Influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: E27813  
R:Kida, H.; Kawaoaka, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: E27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16742; NID:g324091  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: Influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8.22.38.165.285.483/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAFGIENGWEGMIDGMYG 24  
Db 330 GLFGAIAFGIENGWEGMIDGMYG 352

RESULT 9  
HMIY15  
hemagglutinin precursor - Influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: G27813  
R:Kida, H.; Kawaoaka, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: G27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16743; NID:g324093; PIDN:AAA43149.1; PID:g324094  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: Influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Product: hemagglutinin HA2 #status predicted <TM1>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8.22.38.165.285.483/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAFGIENGWEGMIDGMYG 24  
Db 330 GLFGAIAFGIENGWEGMIDGMYG 352

RESULT 10  
JQI153  
hemagglutinin precursor - Influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000  
C:Accession: JQI153  
R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.  
J. Gen. Virol. 72, 2007-2010, 1991  
A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian  
A:Accession: JQI153  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <YAS>  
A:Cross-references: GB:000929; NID:g221279; PIDN:BA00769.1; PID:g221280  
A:Note: the authors translated the codon GGC for residue 218 as Glu  
C:Superfamily: Influenza virus hemagglutinin  
C:Keywords: glycoprotein; homotrimer  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,4e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDGMYG 24

Db 330 GLFGAIGFIENGEGMIDGMYG 352

RESULT 11

J01154

hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000

C:Accession: J01154

R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.

J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3

A:Reference number: J01153; MUID:91341491

A:Accession: J01154

A:Molecule type: genomic RNA

A:Residues: 1-550 <YAS>

A:Cross-references: GB:D00930; NID:9221273; PIDN:BA00770.1; PID:q221274

A:Note: the authors translated the codon GGC for residue 218 as Glu

A:Note: residues 528-532 are not shown in this publication

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: glycoprotein; homotrimer

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,4e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDGMYG 24

Db 330 GLFGAIGFIENGEGMIDGMYG 352

RESULT 12

J01155

hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000

C:Accession: J01155

R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.

J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3

A:Reference number: J01153; MUID:91341491

A:Accession: J01155

A:Molecule type: genomic RNA

A:Residues: 1-550 <YAS>

A:Cross-references: GB:D00931; NID:9221277; PIDN:BA00771.1; PID:q221278

A:Note: The authors translated the codon GGC for residue 218 as Glu, GCC for residue 538

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: glycoprotein; homotrimer

F:1-338/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match

93.7%; Score 134; DB 2; Length 550;

Best Local Similarity 100.0%; Pred. No. 1,4e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDGMYG 24  
Db 330 GLFGAIGFIENGEGMIDGMYG 352

RESULT 13

HM1VH

hemagglutinin precursor - influenza A virus

C:Species: Influenza A virus

C:Date: 28-Feb-1981 #sequence\_revision 28-Feb-1981 #text\_change 22-Oct-1999

C:Accession: A93705; A93233; A04051; A93231; A94441

R:Both, G.W.; Sleight, M.J.

Nucleic Acids Res. 8, 2561-2575, 1980

A:Title: Complete nucleotide sequence of the haemagglutinin gene from a human influ

A:Reference number: A93705; MUID:81053698

A:Accession: A93705

A:Molecule type: genomic RNA

A:Residues: 1-566 <BOT>

A:Cross-references: GB:V01103

A:Experimental source: strain A/NT/60/68/29C

A:Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A

R:Dopheide, T.A.; Ward, C.W.

FEBS Lett. 110, 181-183, 1980

A:Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.

A:Reference number: A91276; MUID:80179105

A:Contents: annotation; disulfide bonds

R:Gethling, M.J.; Bye, J.; Skehel, J.; Waterfield, M.

Nature 287, 301-306, 1980

A:Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes

A:Reference number: A93233; MUID:81030852

A:Accession: A93233

A:Molecule type: genomic RNA

A:Residues: 1-24, 'S', '26', 'D', 28-159, 'G', 161-197, 'T', 199-241, 'V', 243-249 <GET>

A:Experimental source: strain X-31[H3]

C:Superfamily: Influenza virus hemagglutinin

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>

F:30-482/Domain: transmembrane #status predicted <TM>

F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 1,4e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDGMYG 24

Db 346 GLFGAIGFIENGEGMIDGMYG 368

RESULT 14

HM1VH

hemagglutinin precursor - influenza A virus (strain A/ichi/2/68)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 16-Jul-1999

C:Accession: A93231; A04051

R:Verheym, M.; Fang, R.; Min Jou, W.; Devos, R.; Huybrecock, D.; Saman, E.; Flier

Nature 286, 771-776, 1980

A:Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza stra

A:Reference number: A93231; MUID:80254693

A:Accession: A93231

A:Molecule type: genomic RNA

A:Residues: 1-566 <VER>

A:Cross-references: GB:J02090; NID:9324131; PIDN:AAA3178.1; PID:q324132

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>  
F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 1.4e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAAGFIENGEGMIDGMYG 24  
|||||  
DB 346 GLFGAIAAGFIENGEGMIDGMYG 368

## RESULT 15

HMTVHM

hemagglutinin precursor - Influenza A virus (strain A/Mem/102/72)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 31-Mar-2000

C:Accession: A94441; A04051

R:Stleigh, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.

In Structure and Variation in Influenza Virus, Laver, G., and Alr, G., eds., pp.69-79, B

A:Title: The haemagglutinin gene of Influenza A virus: nucleotide sequence analysis of c

A:Reference number: A94441

A:Accession: A94441

A:Molecule type: genomic RNA

A:Residues: 1-566 <SLD>

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>

F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 1.4e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAAGFIENGEGMIDGMYG 24  
|||||  
DB 346 GLFGAIAAGFIENGEGMIDGMYG 368

Search completed: July 1, 2002, 06:28:11  
Job time: 713 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 1, 2002, 06:31:34 ; Search time 25.37 Seconds

(without alignments)  
36,629 Million cell updates/sec

Title: US-09-461-684-3

Perfect score: 143  
Sequence: 1 GCLFGAIGFIENGWEGMTDGMYG 24Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	93.7	550	1 HEMA_IADH1	P12582 Influenza a
2	134	93.7	550	1 HEMA_IADH2	P12583 Influenza a
3	134	93.7	550	1 HEMA_IADH3	P12584 Influenza a
4	134	93.7	550	1 HEMA_IADH4	P12585 Influenza a
5	134	93.7	550	1 HEMA_IADH5	P12586 Influenza a
6	134	93.7	550	1 HEMA_IADH6	P12587 Influenza a
7	134	93.7	550	1 HEMA_IADH7	P12588 Influenza a
8	134	93.7	550	1 HEMA_IADH8	P43257 Influenza a
9	134	93.7	550	1 HEMA_IADH9	P43258 Influenza a
10	134	93.7	550	1 HEMA_IADH10	P43259 Influenza a
11	134	93.7	550	1 HEMA_IADH11	P43260 Influenza a
12	134	93.7	550	1 HEMA_IADH12	P11133 Influenza a
13	134	93.7	550	1 HEMA_IADH13	P11134 Influenza a
14	134	93.7	550	1 HEMA_IADH14	P03437 Influenza a
15	134	93.7	550	1 HEMA_IADH15	P26134 Influenza a
16	134	93.7	550	1 HEMA_IADH16	P03442 Influenza a
17	134	93.7	550	1 HEMA_IADH17	P26138 Influenza a
18	134	93.7	550	1 HEMA_IADH18	P03449 Influenza a
19	134	93.7	550	1 HEMA_IADH19	P03436 Influenza a
20	133	93.0	561	1 HEMA_IADH20	P12581 Influenza a
21	133	93.0	561	1 HEMA_IADH21	P12582 Influenza a
22	132	92.3	565	1 HEMA_IADH22	P17000 Influenza a
23	132	92.3	565	1 HEMA_IADH23	P17001 Influenza a
24	132	92.3	566	1 HEMA_IADH24	P17002 Influenza a
25	132	92.3	566	1 HEMA_IADH25	P03440 Influenza a
26	132	92.3	566	1 HEMA_IADH26	P26139 Influenza a
27	131	91.6	560	1 HEMA_IADH27	P12589 Influenza a
28	131	91.6	560	1 HEMA_IADH28	P03458 Influenza a
29	131	91.6	565	1 HEMA_IADH29	P16994 Influenza a
30	131	91.6	565	1 HEMA_IADH30	P16995 Influenza a
31	131	91.6	565	1 HEMA_IADH31	P16996 Influenza a
32	131	91.6	565	1 HEMA_IADH32	P16997 Influenza a
33	131	91.6	565	1 HEMA_IADH33	P15658 Influenza a

34	131	91.6	565	1 HEMA_IADH34	P16997 Influenza a
35	131	91.6	565	1 HEMA_IADH35	P16998 Influenza a
36	131	91.6	565	1 HEMA_IADH36	P16999 Influenza a
37	131	91.6	565	1 HEMA_IADH37	P08011 Influenza a
38	131	91.6	565	1 HEMA_IADH38	P17001 Influenza a
39	131	91.6	566	1 HEMA_IADH39	P26141 Influenza a
40	131	91.6	570	1 HEMA_IADH40	P26094 Influenza a
41	131	91.6	570	1 HEMA_IADH41	P26095 Influenza a
42	131	91.6	570	1 HEMA_IADH42	P26096 Influenza a
43	131	91.6	570	1 HEMA_IADH43	P26097 Influenza a
44	131	91.6	570	1 HEMA_IADH44	P26098 Influenza a
45	131	91.6	570	1 HEMA_IADH45	P26099 Influenza a

## ALIGNMENTS

RESULT 1  
HEMA\_IADH1 STANDARD: PRT: 550 AA.  
ID HEMA\_IADH1  
AC P12582; Q84021; Q84022;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hokkaido/5/77).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae.  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=11357;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87265458; PubMed=2440178;  
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;  
RT "Antigenic and genetic conservation of H3 Influenza virus in wild  
ducks";  
RL Virology 159:109-119(1987).  
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: M16737; AAA43143.1; -.  
DR PIR: A27813; HMIIV77.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CARBOHYD 8 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 22 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA; 61705 MW; 7E7ACFE716FC969A CRC64;  
Query Match 93.7%; Score 134; DB 1; Length 550.  
Best Local Similarity 100.0%; Pred. No. 1.le-10;



```
CC Influenza virus A and B group; Influenza A viruses;
CC Influenza A virus.
CC NCBI_TaxID=11360;
RN [1]
RP MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

EMBL: M16740; AAA43146.1; -

DR PIR: D27813; HMIY89.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR Prodom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein.

FT NON\_TER 1 1

FT CHAIN 1 328

FT CARBOHYD 330 550

FT CARBOHYD 8 550

FT CARBOHYD 22 22

FT CARBOHYD 38 38

FT CARBOHYD 165 165

FT CARBOHYD 285 285

FT CARBOHYD 483 483

FT CARBOHYD 550 AA; 61664 MW; A16B2CF8CBBD9D0 CRC64;

SEQUENCE

Query Match 93.7%; Score 134; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 1.1e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIGFIENGEGMIDGWC 24

DB 330 GLFGAIGFIENGEGMIDGWC 352

RESULT 5

HEMA\_IADH5 STANDARD: PRT: 550 AA.

ID HEMA\_IADH5

AC P12586; Q84015; Q84016;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;

DE Hemagglutinin HA2 chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/21/82).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group; Influenza A viruses;

OC Influenza A virus.

OX NCBI\_TaxID=11361;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87265458; PubMed=2440178;

RA Kida H., Kawaka Y., Naeve C.W., Webster R.G.;

RT "Antigenic and genetic conservation of H3 influenza virus in wild

RT ducks.";

RL Virology 159:109-119(1987).

```
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

EMBL: M16741; AAA43147.1; -

DR PIR: E27813; HMIY21.

DR HSSP: P03437; SHMG.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR Prodom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein.

FT NON\_TER 1 1

FT CHAIN 1 328

FT CARBOHYD 330 550

FT CARBOHYD 7 7

FT CARBOHYD 8 8

FT CARBOHYD 22 22

FT CARBOHYD 38 38

FT CARBOHYD 165 165

FT CARBOHYD 285 285

FT CARBOHYD 483 483

FT CARBOHYD 550 AA; 61856 MW; 48401C867A15B8C CRC64;

SEQUENCE

Query Match 93.7%; Score 134; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 1.1e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIGFIENGEGMIDGWC 24

DB 330 GLFGAIGFIENGEGMIDGWC 352

RESULT 6

HEMA\_IADH6 STANDARD: PRT: 550 AA.

ID HEMA\_IADH6

AC P12587; Q84017;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;

DE Hemagglutinin HA2 chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/9/85).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group; Influenza A viruses;

OC Influenza A virus.

OX NCBI\_TaxID=11362;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87265458; PubMed=2440178;

RA Kida H., Kawaka Y., Naeve C.W., Webster R.G.;

RT "Antigenic and genetic conservation of H3 influenza virus in wild

RT ducks.";

RL Virology 159:109-119(1987).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M16742; AAA43148.1; -  
DR PIR: F27813; HMTV98.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KM Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 8 8 Y -> N (IN PIR DATA BANK).  
SO SEQUENCE 550 AA; 61711 MW; 67BCD85F44736CFE CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.le-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFLENGMEGMDGMYG 24  
DB 330 GLFGAAGFLENGMEGMDGMYG 352

RESULT 7  
HEMA\_IADH7 STANDARD: PRT; 550 AA.  
ID HEMA\_IADH7  
AC P12588; Q84018; Q89470;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hokkaido/10/85).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=11363;  
RN SEQUENCE FROM N.A.  
RX MEDLINE=87265458; PubMed=2440178;  
RA Kida H., Kawakita Y., Naeve C.W., Webster R.G.;  
RT "Antigenic and genetic conservation of H3 influenza virus in wild  
RT ducks.";  
RT Virology 159:109-119(1987).  
RL -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M16743; AAA43149.1; -  
DR PIR: G27813; HMTV15.

DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KM Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SO SEQUENCE 550 AA; 61761 MW; 6EF81793281D53EB CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.le-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFLENGMEGMDGMYG 24  
DB 330 GLFGAAGFLENGMEGMDGMYG 352

RESULT 8  
HEMA\_IADHK STANDARD: PRT; 550 AA.  
ID HEMA\_IADHK  
AC P43257;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hong Kong/7/75).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=11364;  
RN SEQUENCE FROM N.A.  
RX MEDLINE=91341491; PubMed=1875195;  
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
RT "Molecular evidence for a role of domestic ducks in the introduction  
RT of avian H3 influenza viruses to pigs in southern China, where the  
RT A/Hong Kong/68 (H3N2) strain emerged.";  
RL J. Gen. Virol. 72:2007-2010(1991).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: D00929; BAA00769.1; -  
DR HSP: P03437; 5HKC.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KM Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SO SEQUENCE 550 AA; 61761 MW; 6EF81793281D53EB CRC64;

FT CARBOHYD 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA: 61549 MW: 8646398829FE1BA9 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 1.le-10; Mismatches 0; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GLFGAAGFIENGWEGMIDWYG 24  
DB 330 GLFGAAGFIENGWEGMIDWYG 352

RESULT 9  
HEMA\_IADHL STANDARD: PRT: 550 AA.  
ID HEMA\_IADHL  
AC P43258;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hong Kong/64/76).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=45412;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91341491; Pubmed=1875195;  
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
RT "Molecular evidence for a role of domestic ducks in the introduction  
of avian H3 influenza viruses to pigs in southern China, where the  
A/Hong Kong/68 (H3N2) strain emerged.";  
RT J. Gen. Virol. 72:2007-2010(1991).  
RL A.  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: D00931; BAA00771.1; -  
DR HSSP: P03437; 2HMG.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR Prodom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA: 61718 MW: A351C56789E4BE9A CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.le-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GLFGAAGFIENGWEGMIDWYG 24  
DB 330 GLFGAAGFIENGWEGMIDWYG 352

RESULT 10  
HEMA\_IAGHK STANDARD: PRT: 550 AA.  
ID HEMA\_IAGHK  
AC P43260;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Goose/Hong Kong/10/76).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=45414;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91341491; Pubmed=1875195;  
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
RT "Molecular evidence for a role of domestic ducks in the introduction  
of avian H3 influenza viruses to pigs in southern China, where the  
A/Hong Kong/68 (H3N2) strain emerged.";  
RT J. Gen. Virol. 72:2007-2010(1991).  
RL A.  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: D00930; BAA00770.1; -  
DR HSSP: P03437; 5HMG.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR Prodom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA: 61676 MW: 9A1ED94DA28BACD2 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.le-10; Mismatches 0; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GLFGAAGFIENGWEGMIDWYG 24  
DB 330 GLFGAAGFIENGWEGMIDWYG 352

RESULT 11  
HEMA\_IAGHZ

```
ID HEMA_I2H2 STANDARD: PRT: 550 AA.
AC P1133; 084019; 084020;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1989 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/81/78).
OC Viruses: ssRNA negative-strand viruses: Orthomyxoviridae;
OC Influenza A virus.
OC Influenza A virus.
OX NCBI_TaxID=11497;
RP SEQUENCE FROM N.A.
RX MEDLINE=8101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
in China."
RL Virology 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
EMBL: M19057; AAA43212.1; -
DR PIR: B29971; HMIVS3.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Hemagglutinin; Envelope protein; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61437 MW; 1F2A7E758C53ICE8 CRC64;
```

```
Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred No. 1.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 GLFGAIGFIENGMEGMDWYG 24
DB 330 GLFGAIGFIENGMEGMDWYG 352
RESULT 12
HEMA_I2H3 STANDARD: PRT: 550 AA.
AC P1134; 084025; 084026;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/126/82).
OC Viruses: ssRNA negative-strand viruses: Orthomyxoviridae;
```

```
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11498;
RP SEQUENCE FROM N.A.
RX MEDLINE=8101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
in China."
RL Virology 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
EMBL: M19056; AAA43211.1; ALT_TERM.
DR PIR: A29971; HMIVS2.
DR HSSP: P03437; 2HMC.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Hemagglutinin; Envelope protein; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61580 MW; 991F6DB8C02F24F2 CRC64;
```

```
Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred No. 1.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 GLFGAIGFIENGMEGMDWYG 24
DB 330 GLFGAIGFIENGMEGMDWYG 352
RESULT 13
HEMA_I2A1C STANDARD: PRT: 566 AA.
AC P03437;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Aichi/2/68).
OC Viruses: ssRNA negative-strand viruses: Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=150147;
RP SEQUENCE FROM N.A.
RX MEDLINE=80254693; PubMed=7402351;
RA Verhoeven M., Fang R., Min Jou W., Devos R., Huybrecock D.,
RA Saman E., Fiers W.;
RT "Antigenic drift between the haemagglutinin of the Hong Kong
```



FT STRAND 310 311  
FT STRAND 318 320  
FT STRAND 323 324  
FT STRAND 331 333  
FT STRAND 337 337  
FT STRAND 347 348  
FT STRAND 350 350  
FT STRAND 351 351  
FT STRAND 352 354  
FT STRAND 355 355  
FT STRAND 359 359  
FT STRAND 360 361  
FT STRAND 367 372  
FT STRAND 374 375  
FT STRAND 378 382  
FT STRAND 383 401  
FT STRAND 406 407  
FT STRAND 421 474  
FT STRAND 475 477  
FT STRAND 482 485  
FT HELIX 491 498  
FT TURN 499 500  
FT HELIX 505 515  
SQ SEQUENCE 566 AA: 63415 MW: E395659C23CAFECA CRC64:

Query Match 93.7%: Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%: Pred. No. 1,1e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24  
Db 346 GLFGAIGFIENGWEGMIDGMYG 368

## RESULT 14

HEMA\_IAD3

STANDARD:

PRT: 566 AA.

ID P26134;

AC 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;

DE Hemagglutinin HA2 chain].

GN HA.

OS Influenza A virus (strain A/Duck/Alberta/78/76).

OC Viruses: ssRNA negative-strand viruses: Orthomyxoviridae;

OC Influenza virus A and B group; Influenza A viruses;

OC Influenza A virus.

OX NCB1\_TaxID-11348;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-92114135; PubMed-1731092;

RA Bean W.J., Schnell M., Keltz J., Kawoka Y., Naeye C., Gorman O.,

RT Webster R.G.;

RT "Evolution of the H3 influenza virus hemagglutinin from human and

RT nonhuman hosts."

RL J. Virol. 66:1129-1138(1992).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC -1- CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC C send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M73771; -; NOT\_ANNOTATED\_CDS.

DR HSPP: P03437; SHMG.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
FT SIGNAL 1 16  
FT CHAIN 17 344  
FT CHAIN 346 566  
FT CARBOHYD 23 23  
FT CARBOHYD 24 24  
FT CARBOHYD 38 38  
FT CARBOHYD 54 54  
FT CARBOHYD 181 181  
FT CARBOHYD 301 301  
FT CARBOHYD 499 499  
SQ SEQUENCE 566 AA: 63534 MW: FE19AB6FF94158B9 CRC64:

Query Match 93.7%: Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%: Pred. No. 1,1e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24  
Db 346 GLFGAIGFIENGWEGMIDGMYG 368

## RESULT 15

HEMA\_IAD3

STANDARD:

PRT: 566 AA.

ID P03442;

AC 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;

DE Hemagglutinin HA2 chain].

GN HA.

OS Influenza A virus (strain A/Duck/Ukraine/1/63).

OC Viruses: ssRNA negative-strand viruses: Orthomyxoviridae;

OC Influenza virus A and B group; Influenza A viruses;

OC Influenza A virus.

OX NCB1\_TaxID-11374;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-8202542; PubMed-6169439;

RA Fang R., Min Jou W., Huybreoek D., Devos R., Piers W.;

RT "Complete structure of A/duck/Ukraine/63 influenza hemagglutinin

RT gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza

RT hemagglutinin."

RL Cell 25:315-323(1981).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC -1- CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: V01087; CAA24271.1; -

DR PIR: A04053; HMYDU.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTN12.

DR ProDom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.

FT SIGNAL 1 16



```

FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63530 MW; E70F87F0AE1178F4 CRC64;

```

```

Query Match 93.7%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 GLFGAIGFIENGWEGMIDGMYG 24
   |||||
Db 346 GLFGAIGFIENGWEGMIDGMYG 368

```

Search completed: July 1, 2002, 06:31:35  
 Job time: 672 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 1, 2002, 06:30:57 ; Search time 83.98 Seconds

(Without alignments)  
49,439 Million cell updates/sec

Title: US-09-461-684-3

Sequence: 1 CGLFGATAGTENGMECHIDWCYC 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19:\*

1: sp.\_archaea:\*\n2: sp.\_bacteria:\*\n3: sp.\_fungi:\*\n4: sp.\_human:\*\n5: sp.\_invertebrate:\*\n6: sp.\_mammal:\*\n7: sp.\_mhc:\*\n8: sp.\_organelle:\*\n9: sp.\_phage:\*\n10: sp.\_plant:\*\n11: sp.\_rodent:\*\n12: sp.\_virus:\*\n13: sp.\_vertebrate:\*\n14: sp.\_unclassified:\*\n15: sp.\_virus:\*\n16: sp.\_bacteriap:\*\n17: sp.\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	93.7	566	12	Q9DHG0
2	134	93.7	566	12	Q67125
3	134	93.7	566	12	Q67126
4	134	93.7	566	12	Q67132
5	134	93.7	566	12	Q98052
6	134	93.7	566	12	Q91MA7
7	134	93.0	301	12	Q910M5
8	133	93.0	301	12	Q9DXE3
9	132	92.3	550	12	Q82499
10	132	92.3	550	12	Q82753
11	132	92.3	550	12	Q82498
12	132	92.3	566	12	Q82496
13	132	92.3	571	12	Q03909
14	131	91.6	109	12	Q67050
15	131	91.6	109	12	Q67051
16	131	91.6	109	12	Q67052

17	131	91.6	109	12	Q67053	067053	Influenza a
18	131	91.6	362	12	Q82513	082513	Influenza a
19	131	91.6	362	12	Q82517	082517	Influenza a
20	131	91.6	362	12	Q84174	084174	Influenza a
21	131	91.6	362	12	Q90KD3	090KD3	Influenza a
22	131	91.6	362	12	Q90KD2	090KD2	Influenza a
23	131	91.6	362	12	Q90KD1	090KD1	Influenza a
24	131	91.6	365	12	Q9DL25	09DL25	Influenza a
25	131	91.6	367	12	Q9DL22	09DL22	Influenza a
26	131	91.6	368	12	Q9DL29	09DL29	Influenza a
27	131	91.6	369	12	Q9DL26	09DL26	Influenza a
28	131	91.6	369	12	Q9DL06	09DL06	Influenza a
29	131	91.6	369	12	P87689	P87689	Influenza a
30	131	91.6	371	12	Q9DL24	09DL24	Influenza a
31	131	91.6	371	12	P87685	P87685	Influenza a
32	131	91.6	373	12	Q9DL20	09DL20	Influenza a
33	131	91.6	374	12	Q9DL21	09DL21	Influenza a
34	131	91.6	375	12	Q9DL27	09DL27	Influenza a
35	131	91.6	375	12	Q9DL05	09DL05	Influenza a
36	131	91.6	376	12	Q9DL30	09DL30	Influenza a
37	131	91.6	377	12	Q9E7P6	09E7P6	Influenza a
38	131	91.6	382	12	Q9E7P5	09E7P5	Influenza a
39	131	91.6	408	12	Q9E7P5	09E7P5	Influenza a
40	131	91.6	409	12	Q9G0L5	09G0L5	Influenza a
41	131	91.6	416	12	Q9GJG4	09GJG4	Influenza a
42	131	91.6	429	12	Q9G0L4	09G0L4	Influenza a
43	131	91.6	438	12	Q9G0L3	09G0L3	Influenza a
44	131	91.6	467	12	Q9E312	09E312	Influenza a
45	131	91.6	467	12	Q9E312	09E312	Influenza a

## ALIGNMENTS

RESULT 1  
Q9DHG0 PRELIMINARY; PRT; 566 AA.  
AC Q9DHG0; 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DD HAEMAGGLUTININ PRECURSOR.  
OS Influenza A virus H3N2.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_Taxid=41857;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CLONE 7A (H3N2);  
RA Moshin M.A., Morris S.J., Smith H., Sweet C.;  
RT "Influenza virus-induced apoptosis: a dual role for viral neuraminidase."  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: HAEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HAEMAGGLUTININ FAMILY.  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC EMBL; AJ289703; CAC18525.1; -.  
DR HSSP; P03437; 2V1U.  
DR InterPro; IPR001364; Hemagglutn.  
DR Pfam; PF00509; Hemagglutinin; 1.  
DR PRINTS; PR00329; Hemagglutn12.  
DR PRODOM; PD000225; Hemagglutn; 1.  
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.  
FT SIGNAL 16  
FT SEQUENCE 566 AA; 63356 MW; 0BA681929300F72F CRC64;

Query Match 93.7%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2	GLFGAIAGFIENGWEGMIDGWYG	24
DB	346	GLFGAIAGFIENGWEGMIDGWYG	368
RESULT	2		
ID	067125	PRELIMINARY:	PRT: 566 AA.
AC	067125:		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	HEMAGGLUTININ.		
GN	HA.		
OS	Influenza A virus.		
OC	Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;		
OC	Influenza virus A and B group; Influenza A viruses.		
OX	NCBI_TaxID=11320;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A/SEAL/MA/3911/92;		
RX	MEDLINE=95146951; PubMed=7844533;		
RA	Callan R.J., Early G., Kida H., Hinshaw V.S.;		
RT	"The appearance of H3 influenza viruses in seals."		
RL	J. Gen. Virol. 76:199-203(1995).		
CC	-1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).		
CC	-1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).		
CC	-1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.		
DR	EMBL: L31949; AAA64229.1; "		
DR	HSSP: P03437; 2YTU.		
DR	InterPro: IPR001364; Hemagglutn.		
DR	Planet: PF00509; Hemagglutinin; 1.		
DR	PRINTS: PR00329; HEMAGGLUTN12.		
DR	PRODOM: PD000225; Hemagglutn; 1.		
KW	Envelope protein; Glycoprotein; Hemagglutinin.		
SO	SEQUENCE 566 AA; 63456 MW; AE556302A9EBB99F CRC64;		
Query Match		93.7%; Score 134; DB 12; Length 566;	
Best Local Similarity		100.0%; Pred. No. 2,2e-10;	
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	2	GLFGAIAGFIENGWEGMIDGWYG	24
DB	346	GLFGAIAGFIENGWEGMIDGWYG	368
RESULT	3		
ID	067126	PRELIMINARY:	PRT: 566 AA.
AC	067126:		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	HEMAGGLUTININ.		
GN	HA.		
OS	Influenza A virus.		
OC	Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;		
OC	Influenza virus A and B group; Influenza A viruses.		
OX	NCBI_TaxID=11320;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A/SEAL/MA/3984/92;		
RX	MEDLINE=95146951; PubMed=7844533;		
RA	Callan R.J., Early G., Kida H., Hinshaw V.S.;		
RT	"The appearance of H3 influenza viruses in seals."		
RL	J. Gen. Virol. 76:199-203(1995).		
CC	-1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).		
CC	-1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).		

CC	-1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR	EMBL: U32024; AAA64228.1; -.
DR	HSSP: P03437; 2V1U.
DR	InterPro: IPR001364; Hemagglutn.
DR	Pfam: PF00509; Hemagglutinin; 1.
DR	PRINTS: PR00329; HEMAGGLUTN12.
DR	ProDom: PD000225; Hemagglutn; 1.
KW	Envelope protein; Glycoprotein; Hemagglutinin.
SQ	SEQUENCE 566 AA; 63441 MW; 590576CB4CEE7D08 CRC64;
QY	Query Match 93.7%; Score 134; DB 12; Length 566;
	Best Local Similarity 100.0%; Pred. No. 2.2e-10;
	Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB	2 GLFGAIAGFIENGWEGMIDGWTG 24
	346 GLFGAIAGFIENGWEGMIDGWTG 368
RESULT 4	
ID 067132	PRELIMINARY; PRT; 566 AA.
AC 067132	
DT 01-NOV-1996 (TREMBlrel. 01, Created)	
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)	
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)	
DE HEMAGGLUTININ.	
GN HA.	
OS Influenza A virus (strain A/Aichi/2/68).	
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.	
OX NCBI_TaxID=150147;	
	[1]
	SEQUENCE FROM N.A.
RC STRAIN=A/AICHI/2/68;	
RA Min J.W., Verhoeven M., Fang R.-X., Devos R., Huylebroeck D.,	
RA Fliers W.;	
RA "Shift and drift in influenza viruses.";	
RL (in) Carille M.J., Collins J.F., Moseley B.E. B. (eds.);	
RL SYMPOSIUM OF THE SOCIETY FOR GENERAL MICROBIOLOGY, pp.285-311,	
RL Cambridge University Press, New York (1981).	
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO	
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).	
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS	
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).	
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.	
CC EMBL: M55059; AAA43239.1; -.	
DR HSSP: P03437; IHGE.	
DR InterPro: IPR001364; Hemagglutn.	
DR Pfam: PF00509; Hemagglutinin; 1.	
DR PRINTS: PR00329; HEMAGGLUTN12.	
DR ProDom: PD000225; Hemagglutn; 1.	
KW Envelope protein; Glycoprotein; Hemagglutinin.	
FT CHAIN 1 344	
FT CHAIN 1 346	
FT CHAIN 1 566	
FT CHAIN 1 566	
SQ SEQUENCE 566 AA; 63441 MW; ESD1B97DF96FEECA CRC64;	
QY	2 GLFGAIAGFIENGWEGMIDGWTG 24
	346 GLFGAIAGFIENGWEGMIDGWTG 368
RESULT 5	
ID 098052	PRELIMINARY; PRT; 566 AA.
AC 098052	
DT 01-FEB-1997 (TREMBlrel. 02, Created)	

DT 01-FEB-1997 (TReMBLrel. 02, last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)  
DE HEMAGGLUTININ PRECURSOR (FRAGMENT).  
OS Influenza A virus.  
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID=11320;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=81053698; PubMed=6253883;  
RA Both G.W., Sleight M.J.;  
RT "Complete nucleotide sequence of the haemagglutinin gene from a human  
influenza virus of the Hong Kong subtype.";  
RL Nucleic Acids Res. 8:2561-2573(1980).  
RN (2)  
RP SEQUENCE OF 17-344 FROM N.A.  
RX MEDLINE=81194918; PubMed=6164798;  
RA Sleight M.J., Both G.W., Underwood P.A., Bender V.J.;  
RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza  
subtype: Correlation of amino acid changes with alterations in viral  
antigenicity.";  
RL J. Virol. 37:845-853(1981).  
RN (3)  
RP SEQUENCE OF 17-566 FROM N.A.  
RX MEDLINE=82033276; PubMed=6169843;  
RA Both G.W., Sleight M.J.;  
RT "Conservation and variation in the hemagglutinins of Hong Kong subtype  
influenza viruses during antigenic drift.";  
RL J. Virol. 39:845-853(1981).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: J02135; AAA3189.1; -.  
DR HSSP: P03437; IHGE.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTIN12.  
DR PRODOM: PD000225; Hemagglutn.1.  
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.  
FT SIGNAL 1 16  
FT CHAIN 17 344 POTENTIAL.  
FT CHAIN 346 566 POTENTIAL.  
SQ SEQUENCE 566 AA; 63414 MW; C447FD465BE4FCF9 CRC64;  
  
Query Match 93.7%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 2 GLFGAIGFIENGWEGMIDGMYG 24  
DB 346 GLFGAIGFIENGWEGMIDGMYG 368  
|||||  
RESULT 6  
O91MA7 PRELIMINARY; PRT; 566 AA.  
AC O91MA7;  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)  
DE HEMAGGLUTININ.  
OS Influenza A virus (A/Hong Kong/1/68(H3N2)).  
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=108859;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/HONG KONG/1/68(H3N2);  
RX MEDLINE=21287244; PubMed=11371620;  
RA Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;  
RT Submitted (Oct-2000) to the EMBL/GenBank/DBD databases.

RT "Pattern of mutation in the genome of Influenza A virus on adaptation  
to increased virulence in the mouse lung: Identification of functional  
themes.";  
RT Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).  
DR EMBL: AF348176; AAK51718.1; -.  
SQ SEQUENCE 566 AA; 63387 MW; 01BBD0465BE158E1 CRC64;  
  
Query Match 93.7%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 2 GLFGAIGFIENGWEGMIDGMYG 24  
DB 346 GLFGAIGFIENGWEGMIDGMYG 368  
|||||  
RESULT 7  
O91OM5 PRELIMINARY; PRT; 566 AA.  
AC O91OM5;  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)  
DE HEMAGGLUTININ.  
OS Influenza A virus (A/Hong Kong/1/68(H3N2)).  
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=108859;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/HONG KONG/1/68(H3N2);  
RX MEDLINE=21287244; PubMed=11371620;  
RA Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;  
RT "Pattern of mutation in the genome of influenza A virus on adaptation  
to increased virulence in the mouse lung: Identification of functional  
themes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).  
DR EMBL: AF348177; AAK51721.1; -.  
DR EMBL: AF348177; AAK51719.1; -.  
DR EMBL: AF348178; AAK51720.1; -.  
SQ SEQUENCE 566 AA; 63530 MW; 7CB9F5BAF1E6E9F4 CRC64;  
  
Query Match 93.7%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 2 GLFGAIGFIENGWEGMIDGMYG 24  
DB 346 GLFGAIGFIENGWEGMIDGMYG 368  
|||||  
RESULT 8  
O9DXE3 PRELIMINARY; PRT; 301 AA.  
AC O9DXE3;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)  
DE HEMAGGLUTININ (FRAGMENT).  
GN HA1.  
OS Influenza A virus (A/Shorebird/Taiwan/31-4/99).  
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=140655;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/SHOREBIRD/TAIWAN/31-4/99;  
RA Lee M.S., Cheng P.C., Shien J.H., Cheng M.C., Lee L.H., Shieh H.K.;  
RT "Identification and subtyping of avian influenza virus by reverse  
transcription-polymerase chain reaction.";  
RT Submitted (Oct-2000) to the EMBL/GenBank/DBD databases.

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: AF311750; AAC33016.1; -  
DR InterPro: IPR001364; Hemagglutn.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON\_TER 1 301  
FT CHAIN 1 301  
SQ SEQUENCE 301 AA; 32701 MW; 6244037588764D57 CRC64;

Query Match 93.0%; Score 133; DB 12; Length 301;  
Best Local Similarity 95.7%; Pred. No. 1.5e-10;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLEGAIAGFIEGMEGMDGMYG 24  
DB 250 GLEGAIAGFIEGMEGMDGMYG 272

RESULT 9  
O82499 PRELIMINARY; PRT; 550 AA.  
AC 082499;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE HEMAGGLUTININS HA1 AND HA2 (FRAGMENT).  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID-11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/PHILIPPINES/2/82/BS (H3N2);  
RA Hartley C.A., Ward A.C., Anders E.M.;  
RT "Virulence of influenza virus for mice is associated with loss of  
RT oligosaccharide from the hemagglutinin molecule."  
RL Submitted (Apr-1994) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: U08859; AAA18782.1; -  
DR HSSP: P03437; 2V1U.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON\_TER 1 328  
FT CHAIN 1 328  
FT CHAIN 1 328  
SQ SEQUENCE 550 AA; 61772 MW; 50BD62B6BF11FD8 CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;  
Best Local Similarity 95.7%; Pred. No. 4e-10; Indels 0; Gaps 0;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLEGAIAGFIEGMEGMDGMYG 24  
DB 330 GLEGAIAGFIEGMEGMDGMYG 352

RESULT 10  
O82753 PRELIMINARY; PRT; 550 AA.

AC 082753;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE HAEMAGGLUTININ (FRAGMENT).  
OS Influenza virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC unclassified Orthomyxoviridae.  
OX NCBI\_TaxID-11309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/PHILIPPINES/2/82/BS/ML10 (H3N2);  
RX MEDLINE-97300854; PubMed-9155874;  
RA Hartley C.A., Reading P.C., Ward A.C., Anders E.M.;  
RT "Changes in the hemagglutinin molecule of influenza type A (H3N2)  
RT virus associated with increased virulence for mice."  
RL Arch. Virol. 142:75-88(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/PHILIPPINES/2/82/BS/ML10 (H3N2);  
RX MEDLINE-97456249; PubMed-9311563;  
RA Ward A.C.;  
RT "Virulence of influenza A virus for mouse lung."  
RL Virus Genes 14:187-194(1997).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: U08905; AAC79579.1; -  
DR HSSP: P03437; 2V1U.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON\_TER 1 328  
FT CHAIN 1 328  
FT CHAIN 1 328  
SQ SEQUENCE 550 AA; 61745 MW; 692449DE678AC4BC CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;  
Best Local Similarity 95.7%; Pred. No. 4e-10; Indels 0; Gaps 0;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLEGAIAGFIEGMEGMDGMYG 24  
DB 330 GLEGAIAGFIEGMEGMDGMYG 352

RESULT 11  
O82498 PRELIMINARY; PRT; 550 AA.  
AC 082498;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE HEMAGGLUTININS HA1 AND HA2 (FRAGMENT).  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID-11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/PHILIPPINES/2/82 (H3N2);  
RA Hartley C.A., Ward A.C., Anders E.M.;  
RT "Virulence of influenza virus for mice is associated with loss of  
RT oligosaccharide from the hemagglutinin molecule."  
RL Submitted (Apr-1994) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/PHILIPPINES/2/82 (H3N2);

RX MEDLINE=88185444; PubMed=3356226;  
RA Nakajima S., Takeuchi Y., Nakajima K.:  
RT "Location on the evolutionary tree of influenza H3 haemagglutinin  
RL genes of Japanese strains isolated during the 1985-6 season."  
CC Epidemiol. Infect. 100:301-310(1988).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: U08858; AAA18781.1; -.  
DR HSSP: P03437; 2V1U.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; Hemagglutn12.  
DR PRODOM: PD000225; Hemagglutn.1.  
DR Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HA1.  
FT CHAIN 330 550 HA2.  
SQ SEQUENCE 550 AA; 61802 MW; 11441381C5A1F6A CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;  
Best Local Similarity 95.7%; Pred. No. 4e-10;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GJFGAIAFGIENGWEGMIDGWC 24  
DB 330 GJFGAIAFGIENGWEGMIDGWC 352

RESULT 12  
ID 082496 PRELIMINARY; PRT; 566 AA.  
AC 082496;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEMAGGLUTININ.  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCB1\_TaxID=11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/SWINE/ANGE-GARDIEN/150/90(H3N2);  
RX MEDLINE=95205091; PubMed=7897358;  
RA Bixour M.H., Frost E.H., Deslandes S., Talbot B., Weber J.M.,  
RA Elazhary Y.:  
RT "Recent H3N2 swine influenza virus with haemagglutinin and  
RT nucleoprotein genes similar to 1975 human strains."  
RL J. Gen. Virol. 76:697-703(1995).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: U07146; AAA85781.1; -.  
DR HSSP: P03437; 2V1U.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; Hemagglutn12.  
DR PRODOM: PD000225; Hemagglutn.1.  
DR Envelope protein; Glycoprotein; Hemagglutinin.  
SQ SEQUENCE 566 AA; 63362 MW; 1FB8319A567E2FF CRC64;

Query Match 92.3%; Score 132; DB 12; Length 566;  
Best Local Similarity 95.7%; Pred. No. 4.1e-10;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GJFGAIAFGIENGWEGMIDGWC 24  
DB 346 GJFGAIAFGIENGWEGMIDGWC 368

RESULT 13  
ID 003909 PRELIMINARY; PRT; 571 AA.  
AC 003909;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEMAGGLUTININ PRECURSOR.  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCB1\_TaxID=11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Guo Y., Wang M., Kawaka Y., Gorman O.T., Ito T., Webster R.G.:  
RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
DR EMBL: M65018; AAA43151.1; -.  
DR HSSP: P03437; 2V1U.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR PRODOM: PD000225; Hemagglutn.1.  
DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 349 HA1 CHAIN.  
FT CHAIN 350 571 HA2 CHAIN.  
FT DISULFID 36 487 BY SIMILARITY.  
FT DISULFID 73 298 BY SIMILARITY.  
FT DISULFID 85 97 BY SIMILARITY.  
FT DISULFID 160 494 BY SIMILARITY.  
SQ SEQUENCE 571 AA; 64104 MW; 718DA0F291CE349 CRC64;

Query Match 92.3%; Score 132; DB 12; Length 571;  
Best Local Similarity 95.7%; Pred. No. 4.1e-10;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GJFGAIAFGIENGWEGMIDGWC 24  
DB 351 GJFGAIAFGIENGWEGMIDGWC 373

RESULT 14  
ID 067050 PRELIMINARY; PRT; 109 AA.  
AC 067050;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEMAGGLUTININ (FRAGMENT).  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCB1\_TaxID=11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/FUKUOKA/C29/85 (H3N2);  
RX MEDLINE=81030852; PubMed=7421990;  
RA Gething M.-J., Bye J., Skemel J., Waterfield M.:  
RT "Cloning and dna sequence of double-stranded copies of haemagglutinin  
RT genes from h2 and h3 strains elucidates antigenic shift and drift in  
RL human influenza virus."  
RL Nature 287:301-306(1980).  
[2]

RP SEQUENCE FROM N.A.  
RC STRAIN-A/FUKUOKA/C29/85 (H3N2);  
RX MEDLINE-93233219; PubMed-7682624;  
RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;  
RT "A common neutralizing epitope conserved between the hemagglutinins of  
influenza A virus H1 and H2 strains.";  
RL J. Virol. 67:2552-2558(1993).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC EMBL: D13581; BAA02776.1; -.  
DR HSSP: P03437; 1HTE.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTIN12.  
DR Prodom: PD000225; Hemagglutn.1.  
KM Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON\_TER 1 109 109  
SQ SEQUENCE 109 AA; 12305 MW; 17EC66753C48672F CRC64;

Query Match 91.6%; Score 131; DB 12; Length 109;  
Best Local Similarity 91.3%; Pred. No. 8.9e-11;  
Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFLENGEGMIDGMYG 24  
DB 40 GLFGAIGFLENGEGMIDGMYG 62

RESULT 15  
O67051  
ID 067051 PRELIMINARY; PRT: 109 AA.  
AC 067051;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEMAGGLUTININ (FRAGMENT).  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID=11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/SICHUAN/2/87 (H3N2);  
RX MEDLINE-81030852; PubMed-7421990;  
RA Gething M.-J., Bye J., Skehel J., Waterfield M.;  
RT "Cloning and dna sequence of double-stranded copies of haemagglutinin  
genes from h2 and h3 strains elucidates antigenic shift and drift in  
human influenza virus.";  
RT Nature 287:301-306(1980).  
RL [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/SICHUAN/2/87 (H3N2);  
RX MEDLINE-93233219; PubMed-7682624;  
RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;  
RT "A common neutralizing epitope conserved between the hemagglutinins of  
influenza A virus H1 and H2 strains.";  
RL J. Virol. 67:2552-2558(1993).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC EMBL: D13582; BAA02777.1; -.  
DR HSSP: P03437; 1HTE.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTIN12.

DR Prodom: PD000225; Hemagglutn.1.  
KM Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON\_TER 1 109 109  
SQ SEQUENCE 109 AA; 12293 MW; 17EC66752DB8672F CRC64;

Query Match 91.6%; Score 131; DB 12; Length 109;  
Best Local Similarity 91.3%; Pred. No. 8.9e-11;  
Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFLENGEGMIDGMYG 24  
DB 40 GLFGAIGFLENGEGMIDGMYG 62

Search completed: July 1, 2002, 06:30:58  
Job time: 690 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:23 ; Search time 98.97 Seconds

(Without alignments)  
26.935 Million cell updates/sec

Title: US-09-461-684-3

Sequence: 1 CGLFGAIAAGFTENGMEGNIDGWC 24

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.Geneseq\_032802:\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	100.0	24	21	AA13782
2	143	100.0	26	18	AA26004
3	143	100.0	26	18	AAW34269
4	143	100.0	26	22	AA170090
5	143	100.0	44	21	AA13784
6	143	93.7	23	18	AAW29226
7	134	93.7	25	17	AA198016
8	134	93.7	26	13	AA198005
9	134	93.7	26	18	AAW34270
10	134	93.7	26	22	AA170091
11	134	93.7	28	17	AA198024

12	134	93.7	30	21	AA198680	Core polypeptide f
13	134	93.7	30	22	AB101088	Viral DP178/107-11
14	134	93.7	30	22	AB102537	Viral core polypep
15	134	93.7	30	22	AA13634	DP178-like/DP107-1
16	134	93.7	30	22	AA178081	Core polypeptide T
17	134	93.7	175	15	AA15797	Sequence of BHA2.
18	134	93.7	223	12	AA14032	Peptide sequence f
19	134	93.7	347	15	AA163591	Stem region of A2/
20	134	93.7	566	15	AA163590	Full length H3N2 f
21	134	93.7	685	22	AA17893	Modified clostridi
22	132	92.3	29	20	AA129710	Influenza virus an
23	132	92.3	221	14	AA138864	Sequence of the HA
24	132	92.3	221	14	AA138865	Influenza of the HA
25	132	92.3	221	15	AA160220	Influenza haemagl
26	132	92.3	221	15	AA160221	Influenza haemagl
27	132	92.3	306	14	AA138867	Sequence of fusion
28	132	92.3	306	15	AA160194	Immunogenic fragme
29	132	92.3	386	5	AA140615	Sequence of the X-
30	131	91.6	565	8	AA170711	Equine Influenza v
31	131	91.6	565	11	AA140493	Equine hemagglutin
32	131	91.6	565	19	AA144946	EIV Fontainebleau
33	131	91.6	565	21	AA170056	Wild type equine f
34	131	91.6	565	21	AA170057	Cold-adapted equine
35	131	91.6	566	19	AA168406	SIV strain H3N2 ha
36	131	91.6	570	11	AA170710	Equine Influenza v
37	131	91.6	570	11	AA160490	Influenza A/Beijin
38	131	91.6	570	18	AA160169	Influenza virus A/
39	131	91.6	570	20	AA175441	Influenza virus A/
40	131	91.6	570	22	AA160491	Influenza virus A/
41	131	91.6	571	18	AA160167	Influenza A/Shando
42	131	91.6	571	18	AA160166	Influenza A/Johann
43	131	91.6	571	20	AA175445	Influenza virus A/
44	131	91.6	571	20	AA175448	Influenza virus A/
45	131	91.6	571	22	AA160495	Influenza virus A/

#### ALIGNMENTS

RESULT 1	
ID	AA13782
XX	AA13782 standard; peptide: 24 AA.
XX	
AC	AA13782:
XX	
XX	10-NOV-2000 (first entry)
DT	
XX	
DE	Soluble peptide antigen HA.
KW	HA peptide: cytostatic; vaccine: cytotoxic T cell; CTL; immunotherapy;
KW	major histocompatibility complex class I; MHC class I; antigen; tumour;
KW	prostate; breast; multiple myeloma.
XX	
OS	Unidentified.
XX	
PN	WO200035949-A1.
XX	
PD	22-JUN-2000.
XX	
PF	14-DEC-1999: 99WO-US29724.
XX	
PR	14-DEC-1998: 98US-0112324.
XX	
PA	(DEND-) DENDREON CORP.
XX	
PI	Laus R, Hakim I, Vidovic D;
XX	
DR	WPI: 2000-442365/38.
XX	
PT	Antigens modified by the covalent addition of a peptide that
PT	facilitates entry into antigen presenting cells; useful for producing
PT	compositions for immunizing against tumors and pathogens -



OY 1 CGLFGAAGFIENGWEGMIDGWY 24  
DB 3 cglfgaagfiengwegmidgwy 26

RESULT 4  
AAB70090  
ID AAB70090 standard; peptide: 26 AA.  
XX  
AC AAB70090;  
XX  
DT 14-MAY-2001 (first entry)  
XX  
DE Cationic virosome crosslinker #1 for polypeptide binding.  
XX  
KM Cationic virosome; cytostatic; gene therapy; lipid bilayer vesicle;  
KM viral haemagglutinin; drug delivery; cancer; leukaemia; viral infection;  
KM crosslinker.  
XX  
OS Unidentified.  
XX  
PN NZ504444-A.  
XX  
PD 24-NOV-2000.  
XX  
PF 10-MAY-2000: 2000NZ-0504444.  
XX  
PR 10-MAY-2000: 2000NZ-0332666.  
XX  
PA (NIKA-) NIKA HEALTH PROD LTD.  
XX  
PI Walti ER, Gluck R, Klein P;  
XX  
DR WPI: 2001-233042/24.  
XX  
PT Lipid bilayer vesicle, useful for delivering drugs to target cells such  
PT as cancer, leukemic, or virally infected cells, comprises viral  
PT glycoproteins in positively charged membrane  
XX  
PS Disclosure: Page 14; 41pp: English.  
XX  
CC The present sequence is a crosslinker which is capable of linking to the  
CC surface of a novel cationic virosome and is capable of binding  
CC polypeptides. The cationic virosome is a lipid bilayer vesicle comprising  
CC a membrane with a net positive charge and contains 5-30 weight % based on  
CC total lipids, 1,3-dioleoyloxy-2-(6-carboxy-spermyl)-propyl-amide  
CC (DOSPER), together with other lipids and at least one active fusogenic  
CC peptide. The fusogenic peptide is a viral haemagglutinin that causes the  
CC vesicle to be internalised by target cells through phagocytosis or  
CC endocytosis. The virosome is useful for delivering a desired drug or  
CC substance, preferably a nucleic acid, to target cells (resting or  
CC proliferating mammalian cells) such as cancer cells, leukemic cells or  
CC virally infected cells in vitro, in diagnostic or medical applications  
CC and for the manufacture of medicament for carrying out non-infectious  
CC gene therapy.  
XX  
SQ Sequence 26 AA:

Query Match 100.0%; Score 143; DB 22: Length 26;  
Best Local Similarity 100.0%; Pred. No. 1.4e-14;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAAGFIENGWEGMIDGWY 24  
DB 3 cglfgaagfiengwegmidgwy 26

RESULT 5  
AAB13784  
ID AAB13784 standard; peptide: 44 AA.  
XX  
AC AAB13784;

XX  
DT 10-NOV-2000 (first entry)  
XX  
DE Soluble tandem HA/ pk peptide conjugate.  
XX  
KM pk peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;  
KM major histocompatibility complex class 1; MHC class 1; antigen; tumour;  
KM prostate; breast; multiple myeloma; HA peptide.  
XX  
OS Unidentified.  
XX  
PN WO200035949-A1.  
XX  
PD 22-JUN-2000.  
XX  
PF 14-DEC-1999: 99WO-US29724.  
XX  
PR 14-DEC-1998: 98US-0112324.  
XX  
PA (DEND-) DENDREON CORP.  
XX  
PI Laus R, Hakim I, Vidovic D;  
XX  
DR WPI: 2000-442365/38.  
XX  
PT Antigens modified by the covalent addition of a peptide that  
PT facilitates entry into antigen presenting cells, useful for producing  
PT compositions for immunizing against tumors and pathogens -  
XX  
PS Claim 2; Page 26; 34pp: English.  
XX  
CC The present invention relates to compositions of modified soluble protein  
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)  
CC response i.e. a major histocompatibility complex (MHC) class I molecule  
CC response. The protein antigen is modified by the covalent addition of a  
CC peptide sequence which facilitate entry of the antigen into antigen  
CC presenting cells (APCs). The present sequence is one such peptide  
CC sequence which can be used to modify the soluble antigens. The present  
CC composition may be used for immunising against, or treating a tumour e.g.  
CC prostate and breast carcinoma or multiple myeloma, or pathogen in  
CC mammals.  
XX  
SQ Sequence 44 AA:

Query Match 100.0%; Score 143; DB 21: Length 44;  
Best Local Similarity 100.0%; Pred. No. 2.5e-14;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAAGFIENGWEGMIDGWY 24  
DB 1 cglfgaagfiengwegmidgwy 24

RESULT 6  
AAM29226  
ID AAM29226 standard; peptide: 23 AA.  
XX  
AC AAM29226;  
XX  
DT 02-MAR-1998 (first entry)  
XX  
DE Membrane active acidic peptide INF6 from WO9730170.  
XX  
KM Membrane active; acidic peptide; cationic lipid; transfection;  
KM gene therapy; tumour; vaccine; cytokine gene.  
XX  
OS Synthetic.  
OS Influenza virus.  
XX  
PN WO9730170-A1.

PD 21-AUG-1997.  
XX  
PF 13-FEB-1997; 97WO-EP00649.  
PR 15-FEB-1996; 96DE-4005548.  
XX  
PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.  
XX  
PI Kichler A, Mechtler K, Wagner E;  
XX WPI: 1997-425044/39.  
DR  
XX  
PT Composition containing nucleic acid complexed with cationic lipid  
PT and a membrane-active, acidic peptide - used to increase efficiency  
PT of transfection by destabilising the endosomal membrane, especially  
PT useful for gene therapy  
XX  
PS Claim 4; Page 33; 63pp; German.  
XX  
CC A new composition has been developed for transfecting higher eukaryotic  
CC cells contains: (i) a complex of nucleic acid to be expressed plus a  
CC suboptimal concentration of one or more cationic lipids, and optionally  
CC at least one auxiliary lipid; and (ii) the new feature, at least one  
CC membrane-active, acidic peptide; where the ratio total positive  
CC charges:total negative charges in the composition is about 0.3,  
CC preferably 0-2. The present sequence represents a specifically claimed  
CC acidic peptide for use in the composition. The composition is used for  
CC transfection in vitro, in vivo or ex vivo, particularly for gene  
CC therapy. A typical application is production of tumour vaccines from  
CC autologous cells transfected with a cytokine gene. The acidic peptide  
CC destabilises the endosomal membrane and significantly increases  
CC transfection efficiency, e.g. by as much as 1000 times, equivalent to  
CC that achieved with the optimal amount of cationic lipid. The acidic  
CC peptide also makes the transfection complex less sensitive to serum.  
XX  
SQ Sequence 23 AA;  
  
Query Match 93.7%; Score 134; DB 18; Length 23;  
Best Local Similarity 100.0%; Pred. No. 2.6e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 GLFGAIAAGFLENGEGMIDGWTG 24  
Db 1 glfgaiaagflegngewgidwyg 23  
  
RESULT 7  
AAR98016  
ID AAR98016 standard; peptide; 25 AA.  
XX  
AC AAR98016;  
XX  
DT 12-NOV-1996 (first entry)  
XX  
DE Fusogenic peptide derived from HA-2 spike glycoprotein.  
XX  
KW Fusogenic peptide; spike glycoprotein; enveloped virus; vaccine; EMDPC;  
KW endosome membrane disruption promoting component; cationic polyamine;  
KW multifunctional molecular complex; foodstuff; herbicide; insecticide;  
KW plant growth regulator; miticide; rodenticide; fungicide; parasiticide;  
KW nematocide; immunisation; pathogen; therapy; autoimmune disease;  
KW hyperproliferating disease.  
XX  
XX Synthetic.  
OS  
XX  
PN WO9610038-A1.  
XX  
PD 04-APR-1996.  
XX  
PF 03-SEP-1995; 95WO-US12502.  
XX  
PR 28-SEP-1994; 94US-0314060.

XX  
XX (APOL-) APOLLON INC.  
XX  
XX Boutin RH;  
PI  
XX  
XX WPI: 1996-200887/20.  
DR  
XX  
PT New complexes for nucleic acid transfer to target cells - comprising  
PT a nucleic acid compsn. and a cationic poly:amine with an endosome  
PT membrane disruption component.  
XX  
PS Disclosure; Page 30; 138pp; English.  
XX  
CC AAR98010-R98041 represent fusogenic peptides derived from spike  
CC glycoproteins of enveloped viruses. These fusogenic peptides act as  
CC endosome membrane disruption promoting components (EMDPCs). These  
CC sequences were used in a transfer moiety which makes up part of the  
CC multifunctional molecular complex of the invention. The multifunctional  
CC molecular complex is used for the transfer of a nucleic acid composition  
CC to a target cell, and comprises the nucleic acid composition and a  
CC transfer moiety comprising one or more cationic polyamine components  
CC bound to the nucleic acid composition. There are several different  
CC cationic polyamines that can be used in the complex, one of which  
CC contains one of these sequences linked via a bridging group to the main  
CC body of the cationic polyamine. The products can be used for delivering  
CC therapeutic agents, vaccines, foodstuffs, herbicides, plant growth  
CC regulators, insecticides, miticides, rodenticides, fungicides,  
CC parasitocides or nematocides. They can also be used for immunising an  
CC individual against a pathogen or for treating an autoimmune disease or  
CC hyperproliferating disease. The complexes provide for a high level of  
CC transfection and expression of the nucleic acid molecules in target  
CC cells.  
XX  
SQ Sequence 25 AA;  
  
Query Match 93.7%; Score 134; DB 17; Length 25;  
Best Local Similarity 100.0%; Pred. No. 2.9e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 GLFGAIAAGFLENGEGMIDGWTG 24  
Db 1 glfgaiaagflegngewgidwyg 23  
  
RESULT 8  
AAR26005  
ID AAR26005 standard; peptide; 26 AA.  
XX  
AC AAR26005;  
XX  
DT 26-JAN-1993 (first entry)  
XX  
DE Influenza fusion peptide #2.  
XX  
KW Haemagglutinin; cysteine; liposome.  
XX  
OS Synthetic.  
XX  
PN EP497997-A.  
XX  
PD 12-AUG-1992.  
XX  
PF 02-FEB-1991; 91EP-0101414.  
XX  
PR 02-FEB-1991; 91EP-0101414.  
XX  
PA (NIKA-) NIKA HEALTH PROD LTD.  
XX  
PI Gluck R, Herrmann P, Klein P;  
XX  
DR WPI: 1992-270078/33.  
XX

PT Drug-contg. phospholipid bilayer vesicle with cell-specific  
PT markers on membrane - where markers have at least 90 per cent  
PT biological activity, used as pharmaceuticals against e.g. cancer  
PT and AIDS  
PS Disclosure: Fig 2: 13pp; English.  
XX  
XX The sequences given in AAR26004-19 are influenza fusion peptides,  
CC derived from the influenza virus haemagglutinin gene which were  
CC used for the preparation of synthetic membrane vesicles. The  
CC arrangement of at least one, pref. three cysteine residues at one  
CC end of these peptides has been found useful for the fusion activity,  
CC for the fusion of the liposome to the target cell membrane. The  
CC liposomes produced using these peptides can contain at least one  
CC active drug and can be used to target cells infected with cancer  
CC or AIDS.  
XX  
XX Sequence 26 AA:

Query Match 93.7%; Score 134; DB 13; Length 26;  
Best Local Similarity 100.0%; Pred. No. 3e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGMEGMDGWC 24  
Db 1 glfgaigfiengwegmdgwyg 23

RESULT 9  
AAM34270  
ID AAM34270 standard: peptide; 26 AA.  
XX  
XX AAM34270:  
XX  
XX 14-MAY-1998 (first entry)  
XX  
XX Synthetic lipid vesicle fusion peptide 2.  
DE  
XX  
XX Virosome; vesicle; cationic lipid bilayer; viral fusion peptide;  
KM drug delivery system; membrane; gene therapy; diagnosis; treatment;  
KM cancer; leukaemia; viral infection.  
XX  
XX Synthetic.  
OS  
XX  
XX WO9741834-A1.  
PN  
XX  
XX 13-NOV-1997.  
PD  
XX  
XX 04-MAY-1997; 97WO-EP02268.  
PE  
XX  
XX 08-MAY-1996; 96EP-0107282.  
PR  
XX  
XX (NIKA-) NIKA HEALTH PROD LTD.  
PA  
XX  
XX Glueck R, Klein P, Maelti ER.  
PI  
XX  
XX WPI: 1997-558673/51.  
DR  
XX  
XX  
PT Vesicle with cationic lipid bilayer that includes viral fusion  
PT peptide - used for delivery of genetic material to cells, especially  
PT for gene therapy of cancer, leukaemia and viral infections  
PT  
XX  
XX Disclosure: Page 10; 52pp; English.  
PS  
XX  
XX Peptides AAM34269-W34284 represent novel lipid vesicles with positively  
CC charged lipid bilayer membranes composed of a cationic and/or  
CC polycationic lipid and at least one natural or synthetic viral fusion  
CC peptide integrated in, or covalently linked to, the membrane. Such  
CC peptides are used as drug delivery systems, preferably for  
CC (non-)specific delivery of genetic material to target cells or tissues,  
CC particularly for diagnosis, treatment (especially antisense treatment)  
CC of cancer, leukaemia and viral infections in humans or animals. Genetic

CC material is delivered, without infection, to resting or proliferating  
CC cells, in vitro or in vivo. When the genetic material is an antisense  
CC molecule, it is targeted to mRNA encoding a (proto)oncogene. The  
CC continuous lipid layer does not leak. The peptides do not need to fuse  
CC with, or destabilise, plasma membranes in order to enter the cytoplasm,  
CC since the fusion peptide ensures cell penetration by endocytosis (after  
CC which fusion of the vesicle and endosomal membrane occurs). The genetic  
CC material thus has a greater chance of reaching the nucleus before it is  
CC degraded or expelled. Transfer of the material is 1000-20000 times more  
CC efficient than when using liposomes or conventional virosomes, so  
CC smaller doses can be used, avoiding possible toxicity associated with the  
CC genetic material.  
XX  
XX Sequence 26 AA:

Query Match 93.7%; Score 134; DB 18; Length 26;  
Best Local Similarity 100.0%; Pred. No. 3e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGMEGMDGWC 24  
Db 1 glfgaigfiengwegmdgwyg 23

RESULT 10  
AAB70091  
ID AAB70091 standard: peptide; 26 AA.  
XX  
XX AAB70091:  
XX  
XX 14-MAY-2001 (first entry)  
XX  
XX Cationic virosome crosslinker #2 for polypeptide binding.  
DE  
XX  
XX Cationic virosome; cytostatic; gene therapy; lipid bilayer vesicle;  
KM viral haemagglutinin; drug delivery; cancer; leukaemia; viral infection;  
KM crosslinker.  
XX  
XX Unidentified.  
OS  
XX  
XX NZ504444-A.  
PN  
XX  
XX 24-NOV-2000.  
PD  
XX  
XX 10-MAY-2000; 2000NZ-0504444.  
PE  
XX  
XX 10-MAY-2000; 2000NZ-0332666.  
PR  
XX  
XX (NIKA-) NIKA HEALTH PROD LTD.  
PA  
XX  
XX Walti ER, Glueck R, Klein P;  
PI  
XX  
XX WPI: 2001-233042/24.  
DR  
XX  
XX  
PT Lipid bilayer vesicle, useful for delivering drugs to target cells such  
PT as cancer, leukemic, or virally infected cells, comprises viral  
PT glycoproteins in positively charged membrane  
PT  
XX  
XX Disclosure: Page 14; 41pp; English.  
PS  
XX  
XX The present sequence is a crosslinker which is capable of linking to the  
CC surface of a novel cationic virosome and is capable of binding  
CC polypeptides. The cationic virosome is a lipid bilayer vesicle comprising  
CC a membrane with a net positive charge and contains 5-30 weight % based on  
CC total lipids, 1,3-dioleoyloxy-2-(6-carboxy-spermyl)-propyl-amide  
CC (DOSEPR), together with other lipids and at least one active fusogenic  
CC peptide. The fusogenic peptide is a viral haemagglutinin that causes the  
CC vesicle to be internalised by target cells through phagocytosis or  
CC endocytosis. The virosome is useful for delivering a desired drug or  
CC substance, preferably a nucleic acid, to target cells (resting or  
CC proliferating mammalian cells) such as cancer cells, leukaemic cells or  
CC virally infected cells in vitro, in diagnostic or medical applications

CC and for the manufacture of medicament for carrying out non-infectious  
CC gene therapy.  
XX  
SQ Sequence 26 AA:

Query Match 93.7%; Score 134; DB 22; Length 26;  
Best Local Similarity 100.0%; Pred. No. 3e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GLFGAAGFLENGMEGMDGMYG 24  
Db 1 glfgaagflegngwgmldgwyg 23

## RESULT 11

AAR98024

ID AAR98024 standard; peptide: 28 AA.

AC AAR98024;

DT 12-NOV-1996 (first entry)

DE Fusogenic peptide derived from HA-2 X:31 spike glycoprotein.

KW Fusogenic peptide; spike glycoprotein; enveloped virus; vaccine; EMDPC;  
KW endosome membrane disruption promoting component; cationic polyamine;  
KW multifunctional molecular complex; foodstuff; herbicide; insecticide;  
KW plant growth regulator; miticide; rodenticide; fungicide; parasiticide;  
KW nematocide; immunisation; pathogen; therapy; autoimmune disease;  
KW hyperproliferating disease.

OS Synthetic.

PN WO9610038-A1.

PD 04-APR-1996.

PE 28-SEP-1995; 95WO-US12502.

PR 28-SEP-1994; 94US-0314060.

PA (APOL-) APOLLON INC.

PI Boutin RH;

DR WPI; 1996-200887/20.

XX New complexes for nucleic acid transfer to target cells - comprising  
PT a nucleic acid compsn. and a cationic poly(amine with an endosome  
PI membrane disruption component.

PS Disclosure; Page 30; 138pp; English.

CC AAR98010-988041 represent fusogenic peptides derived from spike  
CC glycoproteins of enveloped viruses. These fusogenic peptides act as  
CC endosome membrane disruption promoting components (EMDPCs). These  
CC sequences were used in a transfer moiety which makes up part of the  
CC multifunctional molecular complex of the invention. The multifunctional  
CC molecular complex is used for the transfer of a nucleic acid composition  
CC to a target cell, and comprises the nucleic acid composition and a  
CC transfer moiety comprising one or more cationic polyamine components  
CC bound to the nucleic acid composition. There are several different  
CC cationic polyamines that can be used in the complex, one of which  
CC contains one of these sequences linked via a bridging group to the main  
CC body of the cationic polyamine. The products can be used for delivering  
CC therapeutic agents, vaccines, foodstuffs, herbicides, plant growth  
CC regulators, insecticides, miticides, rodenticides, fungicides,  
CC parasiticides or nematocides. They can also be used for immunising an  
CC individual against a pathogen or for treating an autoimmune disease or  
CC hyperproliferating disease. The complexes provide for a high level of  
CC transfection and expression of the nucleic acid molecules in target  
CC cells.

XX  
SQ Sequence 28 AA:

Query Match 93.7%; Score 134; DB 17; Length 28;  
Best Local Similarity 100.0%; Pred. No. 3.3e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GLFGAAGFLENGMEGMDGMYG 24  
Db 1 glfgaagflegngwgmldgwyg 23

## RESULT 12

AA989680

ID AA989680 standard; peptide: 30 AA.

AC AA989680;

DT 23-MAY-2000 (first entry)

DE Core polypeptide fragment T No. 1242.

KW Retrovirus; hybrid polypeptide; enhancer; gp41; envelope protein; HIV-1;  
KW HIV-2; SIV; pharmacokinetic; half-life; growth factor; cytokine; viral;  
KW anti-fusogenic; differentiation factor; interleukin; interferon;  
KW colony stimulating factor; hormone; angiogenic factor.

OS Unidentified.

PN WO959615-A1.

PD 25-NOV-1999.

PE 20-MAY-1999; 99WO-US11219.

PR 20-MAY-1998; 98US-0082279.

PA (TRIM-) TRIMERIS INC.

PI Barney S, Guthrie KI, Merutka G, Anwer MK, Lambert DM;

DR WPI; 2000-136792/12.

XX A new hybrid polypeptide with enhanced pharmacokinetic properties  
PT comprises enhancer sequence -

PS Disclosure; Page 42; 124pp; English.

CC The invention relates to hybrid polypeptides comprising enhancer peptide  
CC sequence linked to core polypeptides. The enhancer polypeptides are  
CC derived from various retroviral envelope (gp41) protein sequences, and  
CC especially from HIV-1, HIV-2 and SIV. The enhancer peptides enhance the  
CC pharmacokinetic properties such as increasing the half-life of any core  
CC polypeptide that they are linked to. The core polypeptides are any  
CC polypeptide that may be introduced into a living system and that can  
CC function as a pharmacologically useful peptide for the treatment or  
CC prevention of a disease. The core polypeptides are bioactive peptides  
CC selected from a growth factor, cytokine, differentiation factor,  
CC interleukin, interferon, colony stimulating factor, hormone or  
CC angiogenic factor. The peptides of the invention can be used for  
CC inhibiting viral infection and can be used in anti-viral and  
CC anti-fusogenic treatments. Sequences AA989651-Y90055 represent core  
CC polypeptide fragments that can be used in the invention. Some sequences  
CC among those indicated also comprise enhancer fragments at terminal ends  
CC and form hybrid polypeptides.

SQ Sequence 30 AA;

Query Match 93.7%; Score 134; DB 21; Length 30;  
Best Local Similarity 100.0%; Pred. No. 3.6e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```

OS Synthetic.
XX
XX Key Location/Qualifiers
FH Modified-size 1 /note- "N-terminal is substituted by Ac"
FT Modified-size 30
FT Modified-size 30 /note- "C-terminal amide"
XX
XX WO200151673-A2.
XX
XX 19-JUL-2001.
XX
XX 05-JUL-2000; 2000WO-US35727.
XX
XX 09-JUL-1999; 99US-0350841.
XX
XX (TRIM-) TRIMERIS INC.
XX
XX Jeffs P, Lackey JW, Erickson JB, Lawless MK, Merutka G;
XX
XX WPI; 2001-442157/47.
XX
XX
XX PT Identifying a compound that inhibits the formation of or disrupts a
XX DP107/DP178 complex, especially compounds with antifusogenic, antiviral
XX or intracellular modulatory activity, by detecting the formation of a
XX DP107/DP178 complex -
XX
XX PS Disclosure; Page 74; 259pp; English.
XX
XX CC The present invention relates to peptides which exhibit anti-retroviral
XX activity. The peptides of the invention (AAU12559-AAU14009) comprise
XX DP178-like and DP107-like peptides. The DP178 peptide corresponds
XX to amino acids 639-673 of the transmembrane protein gp41 from human
XX immunodeficiency virus 1 (HIV-1) isolate LAI. The DP107 peptide
XX corresponds to amino acids 558-595 of gp41 from HIV-1LAI. The invention
XX also relates to a method of identifying compounds that inhibit the
XX formation of or disrupts a DP107/DP178 complex. The method comprises
XX detecting the formation of a DP107/DP178 complex, both in the presence
XX or absence of a test compound, in a reaction mixture containing DP107
XX and DP178 peptides. The method is useful for identifying compounds,
XX including small molecule compounds, which may themselves exhibit
XX antifusogenic, antiviral or intracellular modulatory activity. The
XX DP178-like/DP107-like peptides are useful to inhibit human and non-human
XX retroviral, particularly HIV, transmission to uninfected cells. The
XX present sequence represents one of the DP178-like/DP107-like peptides
XX of the invention.
XX
XX SQ Sequence 30 AA:

```

Query Match 93.7%; Score 134; DB 22; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 3, 6e-13;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 2 GLFGAINGFIENGWEGMIDGWY 24
   |||
Db 1 gllfgaiaqlengwemldgwyg 23

```

Search completed: July 1, 2002, 06:19:24  
 Job time: 511 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 1, 2002, 06:28:11 ; Search time 49.45 Seconds

(Without alignments)  
87,442 Million cell updates/sec

Title: US-09-461-684-4

Perfect score: 209  
Sequence: 1 CEAAAAAAAAAAAAAAAAE.....KKKKKKKKKKKKKKKKKK 45Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

PIR-71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103	49.3	248	1	HSUR1P
2	102	48.8	517	2	T49173
3	101.5	48.6	442	2	T39683
4	100	47.8	215	2	I52523
5	100	47.8	380	2	T46395
6	98	46.9	206	2	S09388
7	96	45.9	433	2	S25194
8	95	45.5	166	2	T18513
9	91	43.5	229	2	UC7219
10	90.5	43.3	383	2	AB6315
11	90	43.1	153	2	S59591
12	90	43.1	483	2	F71619
13	90	43.1	529	2	T50609
14	89	42.6	208	2	T23778
15	89	42.6	265	2	S19113
16	88	42.1	392	2	T15755
17	88	42.1	409	2	T18726
18	87	41.6	107	2	C86477
19	87	41.6	565	2	T47775
20	86	41.1	347	2	E83525
21	86	41.1	455	2	A87913
22	85.5	40.9	1280	2	G96796
23	85	40.7	102	2	D64363
24	85	40.7	153	2	S59587
25	85	40.7	241	2	JN0748
26	85	40.7	560	2	T06377
27	85	40.7	849	2	T01286
28	84.5	40.4	228	2	T36379
29	84	40.2	409	2	T24543

30	84	40.2	425	2	T18723	hypothetical prote
31	84	40.2	441	2	A48455	acidic phosphoprot
32	84	40.2	1560	2	T42727	proliferation pote
33	84	40.2	1701	2	T09127	probable erythrocy
34	83	39.7	107	2	A75031	lsu ribosomal prot
35	83	39.7	146	2	A13340	hypothetical prote
36	83	39.7	320	2	S61586	probable membrane
37	83	39.7	421	2	JV0057	toIA protein - Esc
38	83	39.7	625	2	T39019	probable mitochond
39	82	39.2	153	2	S59125	histone H2B [valid
40	82	39.2	474	2	T38485	centromere/microt
41	81.5	39.0	628	2	T08942	proton pump intera
42	81	38.8	111	2	F71216	probable ribosomal
43	81	38.8	218	2	JC7220	nuclear protein SR
44	80.5	38.5	1002	2	S70292	FUN12 protein - ye
45	80	38.3	111	2	E72524	probable ribosomal

## ALIGNMENTS

```
RESULT 1
HSUR1P
histone H1, gonadal - sea urchin (Parechinus angulosus)
C:Species: Parechinus angulosus (angulate urchin)
C:Date: 31-Mar-1980 #sequence_revision 31-Mar-1980 #text_change 16-Feb-1997
C:Accession: A91090; A91091; A02586
R:Strickland, W.N.; Strickland, M.; de Groot, P.C.; von Holt, C.; Wittmann-Liebold, B
Eur. J. Biochem. 104, 559-566, 1980
A:Title: The primary structure of histone H1 from sperm of the sea urchin Parechinus
fimal cyanogen bromide peptides.
A:Reference number: A91090; MUID:80156831
A:Contents: sequence of residues 1-84
A:Accession: A91090
A:Molecule type: protein
A:Residues: 1-248 <STR>
R:Strickland, W.N.; Strickland, M.; Brandt, W.F.; von Holt, C.; Lehmann, A.; Wittmann
Eur. J. Biochem. 104, 567-578, 1980
A:Title: The primary structure of histone H1 from sperm of the sea urchin Parechinus
A:Reference number: A91091; MUID:80156832
A:Accession: A91091
A:Molecule type: protein
A:Residues: 80-248 <ST2>
A:Note: 144-Arg was also found
C:Superfamily: histone H1
C:Keywords: DNA binding; nucleosome; sperm

Query Match 49.3% Score 103. DB 1: Length 248:
Best Local Similarity 59.5% Pred. No. 0.087;
Matches 25; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

QY 2 EAAAAAAAAAAAAAAAAAKKKKKKKKKKKKKKKKK 43
Db 146 KAAAKRAALAKKKAALAKKKAALAKKKAALAKK 187

RESULT 2
T49173
hypothetical protein T20N10.250 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000
R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.
submitted to the Protein Sequence Database, April 2000
A:Reference number: 225017
A:Accession: T49173
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-517 <DAN>
A:Cross-References: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250
A:Experimental source: cultivar Columbia; BAC clone T20N10
C:Genetics:
```

A:Gene: ATSP:T20N10.250  
A:Map position: 3  
A:Introns: 312/3: 359/3: 444/3  
C:Superfamily: Arabidopsis thaliana hypothetical protein F1J16.30

Query Match 48.8%; Score 102; DB 2; Length 517;  
Best Local Similarity 80.8%; Pred. No. 0.17;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 20 EAAAAA KKKKKKKKKKKKKKKKKKKKK 45  
DB 440 ERVGKKKKKKKKKKKKKKKKKKKKKK 465

RESULT 3  
T39683  
zuotin-like protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000  
C:Accession: T39683; T40195  
R:Oliver, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, March 1998  
A:Reference number: Z21869  
A:Accession: T39683  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-124 <OL>  
A:Cross-references: EMBL:AL049489; PIDN:GAB39796.1; GSPDB:GN00067; SPDB:SPBC1778.01c  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.  
submitted to the EMBL Data Library, February 1998  
A:Reference number: Z21910  
A:Accession: T40195  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 89-442 <MOO>  
A:Cross-references: EMBL:Z97992; PIDN:GAB10796.1; GSPDB:GN00067; SPDB:SPBC30D10.01  
A:Experimental source: strain 97zh-; cosmid c30D10  
C:Genetics:  
A:Gene: SPDB:SPBC1778.01c; SPDB:SPBC30D10.01  
A:Map position: 2

Query Match 48.6%; Score 101.5; DB 2; Length 442;  
Best Local Similarity 56.0%; Pred. No. 0.17;  
Matches 28; Conservative 3; Mismatches 12; Indels 7; Gaps 1;  
QY 2 EAAAAA-----EAAAAAEAAAAA KKKKKKKKKKKKKKKKKKKKK 44  
DB 309 EAAAAA QKKKKKEERAAAEAAKASAAANKKAKEDKKKAKRDKKVK 358

RESULT 4  
I52523  
nucleoporin p62 homolog - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 27-Feb-1997 #sequence\_revision 27-Feb-1997 #text\_change 05-Nov-1999  
C:Accession: I52523  
R:Wang, Z.Q.; Akmal, K.M.; Kim, K.H.  
Biol. Reprod. 51, 1022-1030, 1994  
A:Title: An unusual nucleoporin-related messenger ribonucleic acid is present in the ger  
A:Reference number: I52523; MUID:95151924  
A:Accession: I52523  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-215 <RES>  
A:Cross-references: GB:S75997; NID:9913245; PIDN:AAB33364.1; PID:9913246  
A:Experimental source: testis

Query Match 47.8%; Score 100; DB 2; Length 215;  
Best Local Similarity 100.0%; Pred. No. 0.13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 26 KKKKKKKKKKKKKKKKKKKKKKKKKKK 45  
DB 35 KKKKKKKKKKKKKKKKKKKKKKKKKKK 54

RESULT 5  
T46395  
hypothetical protein DKFZp434I1120.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T46395  
R:Ottewill, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Mleemann, S.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23031  
A:Accession: T46395  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-380 <AAA>  
A:Cross-references: EMBL:AL137556  
A:Experimental source: adult testis; clone DKFZp434I1120  
C:Genetics:  
A:Note: DKFZp434I1120.1

Query Match 47.8%; Score 100; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 26 KKKKKKKKKKKKKKKKKKKKKKKKKKK 45  
DB 355 KKKKKKKKKKKKKKKKKKKKKKKKKKK 374

RESULT 6  
S09388  
histone H1 - sea urchin (Parechinus angulosus)  
C:Species: Parechinus angulosus (angulate urchin)  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 24-Jul-1997  
C:Accession: S09388  
R:Hill, C.S.; Martin, S.R.; Thomas, J.O.  
EMBO J. 8, 2591-2599, 1989  
A:Title: A stable alpha-helical element in the carboxy-terminal domain of free and c  
A:Reference number: S09388; MUID:90060019  
A:Accession: S09388  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-206 <HIL>  
C:Superfamily: histone H1  
C:Keywords: chromosomal protein

Query Match 46.9%; Score 98; DB 2; Length 206;  
Best Local Similarity 55.8%; Pred. No. 0.19;  
Matches 24; Conservative 4; Mismatches 15; Indels 0; Gaps 0;  
QY 2 EAAAAAEAAAAAEAAAAA KKKKKKKKKKKKKKKKKKKKK 44  
DB 140 KAAAKKRAALAKKRAAAAKKRAATKAKKAPKKTAKKAKK 182

RESULT 7  
S25194  
zuotin - yeast (Saccharomyces cerevisiae)  
N:Alternate names: probable 2-DNA-binding protein; protein G9554; protein YGR285c  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-May-1993 #sequence\_revision 28-May-1993 #text\_change 21-Jul-2000  
C:Accession: S25194; S64620; S19066  
R:Zhang, S.; Lockshin, C.; Herbert, A.; Winter, E.; Rich, A.  
EMBO J. 11, 3787-3796, 1992  
A:Title: Zuotin, a putative 2-DNA binding protein in Saccharomyces cerevisiae.  
A:Reference number: S25194; MUID:93010971





Mon Jul 1 13:55:04 2002

us-09-461-684-4.rpr

Page 5

```

Qy      3  AAAAAEAAAAAAAAEAAAAAKKKKKKKK 34
          || | | | | | | | | | | | | | | |
Db      176 AAEEARAAAAEAAAAAKARAAAAEAKAKADK 207

```

Search completed: July 1, 2002, 06:28:12  
Job time: 714 sec

OY 2 EAAAAAAAAEAAAAAEEAAAAKKKKKKKKKKKKKKK 43

Db 146 KAAAKRKAALAKKKAALAKKKAALAKKKAALAKKKAALAKK 187

```
RESULT 2
ID 2U01_YEAST STANDARD: PRT: 433 AA.
AC P32527;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ZuoLin.
GN 2U01 OR YGR285C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;

RN 111
SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC STRAIN=20B-12;
RA MEDLINE=93010971; PubMed=1396572;
RA Zhang S., Lockshin C., Herbert A., Winter E., Rich A.;
RT "ZuoLin, a putative Z-DNA binding protein in Saccharomyces
RT cerevisiae."
RL EMBL J. 11:3787-3796(1992).

RN 121
SEQUENCE FROM N.A.
RC STRAIN=5288C / FY1679;
RX MEDLINE=97245295; PubMed=9090054;
RA Volckaert G., Voet M., Robben J.;
RT "Sequence analysis of a near-subtelomeric 35.4 kb DNA segment on the
RT right arm of chromosome VII from Saccharomyces cerevisiae carrying
RT the MAL1 locus reveals 15 complete open reading frames, including
RT 2U01, Bgl2 and B102 genes and an ABC transporter gene."
RL Yeast 13:251-259(1997).

CC -1- FUNCTION: Z-DNA BINDING PROTEIN. COULD BE INVOLVED IN
CC CHROMOSOME ORGANIZATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X63612; CAA45156.1; -
DR EMBL: 273070; CAA97317.1; -
DR PIR: S25194; S25194.
DR HSSP: P08622; 1XB1.
DR SGD: S0003517; 2U01.
DR InterPro: IPR001623; DnaJ_N.
DR Pfam: PF00226; DnaJ; 1.
DR SMART: SM00271; DnaJ; 1.
DR PROSITE: PS00636; DnaJ_1; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
KW Chapterone; DNA-binding; Nuclear protein.
KW DOMAIN 98 170 J-DOMAIN.
FT DOMAIN 306 357 ALA/LYS-RICH.
FT SEQUENCE 433 AA; 49019 MW; 0AA76BC1D3C7DAB CRC64;
```

Query Match 45.9%; Score 96; DB 1; Length 433;  
Best Local Similarity 52.3%; Pred. No. 0.16;  
Matches 23; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

OY 2 EAAAAEAAAAEAAAAEAAAAEAAAAEAAAAEAAAAEAAAAEAAAAE  
313 EAAKAEAEAKAESEAKANASAKADKKKAKAEAAKAKAKKKNR 356

RESULT 3

Db H2B4\_CHLRE STANDARD: PRT: 153 AA.

```
AC P54347;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Histone H2B-IV.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;

RN 111
SEQUENCE FROM N.A.
RX MEDLINE=96120862; PubMed=8590479;
RA Fabry S., Mueller K., Lindauer A., Park P.B., Cornelius T.,
RA Schmitz R.;
RT "The organization structure and regulatory elements of Chlamydomonas
RT histone genes reveal features linking plant and animal genes."
RL Curr. Genet. 28:333-345(1995).

CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4: WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U16726; AAA98454.1; -
DR InterPro: IPR000558; Histone_H2B.
DR InterPro: IPR000166; Histone-core.
DR Pfam: PF00125; Histone; 1.
DR PRINTS: PR00621; HISTONEH2B.
DR PRODOM: PD000497; Histone_H2B; 1.
DR SMART: SM00427; H2B; 1.
DR PROSITE: PS00357; HISTONE_H2B; 1.
KW Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding;
KW Multigene family.
SO SEQUENCE 153 AA; 16587 MW; B09C1EDB04E1FA85 CRC64;
```

Query Match 43.1%; Score 90; DB 1; Length 153;  
Best Local Similarity 51.0%; Pred. No. 0.2;  
Matches 25; Conservative 2; Mismatches 16; Indels 6; Gaps 1;

OY 3 AAAAAEAAAAEAAAAEAAAAEAAAAEAAAAEAAAAEAAAAEAAAAE  
12 AENGAEAPAKAEAKPKAKKAKKAKKAKKAKKAKKAKKAKKAKK 60

```
RESULT 4
ID H1L_CAEEL STANDARD: PRT: 207 AA.
AC P10771;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;

RN 111
SEQUENCE FROM N.A.
RX MEDLINE=90204554; PubMed=1969492;
RA Santicola M., Ward S., Childs G., Emmons S.W.;
RT "Identification of a Caenorhabditis elegans histone H1 gene family.
RT Characterization of a family member containing an intron and encoding
```









KW Bromodomain; Repeat; Nuclear protein.  
FT DOMAIN 56 115 BROMODOMAIN 1.  
FT DOMAIN 326 398 BROMODOMAIN 2.  
FT DOMAIN 487 555 LYS-RICH.  
FT DOMAIN 676 725 SER-RICH.  
FT CONFLICT 465 466 EL->DV (IN REF. 2).  
SQ SEQUENCE 726 AA; 79541 MW; 64f526fc3c1033aa CRC64.

Query Match 40.2%; Score 84; DB 1; Length 726;  
Best Local Similarity 54.3%; Pred. No. 2;  
Matches 19; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

OY 11 AAAAAAAAAAAAAAAAAAKKKKKKKKKKKKKKK 45  
DB 472 AVHEQLALSLQAPVKNPKKKKKKKKKKKKK 506

RESULT 11  
ID RL12\_PYRAB STANDARD; PRT; 107 AA.  
AC Q9UXS6;

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 50S ribosomal protein l12p.  
GN RPL12P OR PAB1168.  
OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=29292;

RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN=ORSAY;

RA Hellig R.;  
RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome structure and evolution."  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR ACCURATE TRANSLATION (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AJ248288; CAB50687.1; -  
DR InterPro; IPR001813; 60S\_Ribosomal.  
DR Pfam; PF00428; 60S\_Ribosomal; 1.

KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 107 AA; 11300 MW; CD4239930BAD4ED CRC64;

Query Match 39.7%; Score 83; DB 1; Length 107;  
Best Local Similarity 44.2%; Pred. No. 0.54;  
Matches 19; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAKKKKKKKKKKKKKKK 44  
DB 55 EKAAAPVAAPAAAPAAPEAEKEKEKEKEKEE 97

RESULT 12  
ID YD33\_YEAST STANDARD; PRT; 320 AA.  
AC Q12117;

DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical 36.2 kDa protein in RAD28-LYS14 intergenic region.  
GN YDR033W OR YD9673.03.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN 11

RP SEQUENCE FROM N.A.  
RA Arnold W., Becker A., Jaeger W., Kuester H., Nussbaumer B.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN 12

RC STRAIN=S288C / AB972;  
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

CC -1- SIMILARITY: BELONGS TO THE ARCHAEL OPSIN FAMILY. HSP30 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z74329; CA99855.1; -

DR EMBL; Z68196; CA92370.1; -

DR SGD; S0002440; YDR033W.

DR InterPro; IPR001425; Bac\_rhodopsin.

DR Pfam; PF01036; Bac\_rhodopsin; 1.

KW Hypothetical protein; Transmembrane.

FT TRANSMEM 35 55 POTENTIAL.

FT TRANSMEM 63 83 POTENTIAL.

FT TRANSMEM 117 137 POTENTIAL.

FT TRANSMEM 142 162 POTENTIAL.

FT TRANSMEM 168 188 POTENTIAL.

FT TRANSMEM 205 225 POTENTIAL.

FT TRANSMEM 239 259 POTENTIAL.

FT DOMAIN 300 318 LYS-RICH.

SQ SEQUENCE 320 AA; 36190 MW; 4311f64e6aa209f CRC64;

Query Match 39.7%; Score 83; DB 1; Length 320;  
Best Local Similarity 48.8%; Pred. No. 1.3;  
Matches 21; Conservative 4; Mismatches 14; Indels 4; Gaps 1;

OY 7 AAAAAAAAAAAAAA-----AKKKKKKKKKKKKKKK 45  
DB 277 AEAVMAPKAPVAPRPATPNLSKKKKKKKKSKSK 319

RESULT 13  
ID TOLA\_ECOLI STANDARD; PRT; 421 AA.  
AC P19934;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE TOLA protein.

GN TOLA OR CIM OR EXCC OR LKY OR B0739.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia

OX NCBI\_TaxID=562;

RN 11

RP SEQUENCE FROM N.A.

RC STRAIN=JM105;

RX MEDLINE=90078104; PubMed=2687247;

RA Levene S.K., Webster R.E.;

RT "Nucleotide sequences of the *tola* and *tolB* genes and localization of their products, components of a multistep translocation system in





Gencore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:30:58 ; Search time 83.98 Seconds  
(Without alignments)  
92.698 Million cell updates/sec

Title: US-09-461-684-4  
Perfect score: 209  
Sequence: 1 CEAAAAAAAAAAAAAAAAE.....KKKKKKKKKKKKKKKKKK 45

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMBL\_19:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*  
16: sp\_rv1rus:\*  
17: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118	56.5	2451	5	09VG05
2	105	50.2	129	11	035807
3	105	50.2	720	4	09H607
4	102	48.6	467	10	09L62
5	102	48.6	517	10	09LXR2
6	101.5	48.6	354	3	014347
7	101	48.3	531	6	0951V6
8	100	47.8	128	3	09P529
9	100	47.8	168	4	09H5V6
10	100	47.8	215	11	064075
11	100	47.8	260	10	09L629
12	100	47.8	380	4	09NT34
13	99.5	47.6	377	5	09GN12
14	99	47.4	686	4	09NXP0
15	97	46.4	667	4	09HC48
16	95	45.5	257	4	09H5Y3

17	92.5	44.3	372	2	09WXX1	09WXX1 pseudomonas
18	91	43.5	229	11	09JMG3	09JMG3 mus musculus
19	91	43.5	263	4	09B076	09B076 homo sapien
20	90.5	43.3	368	10	0944L9	0944L9 arabidopsis
21	90.5	43.3	383	10	09LMT3	09LMT3 arabidopsis
22	90	43.1	166	5	099281	099281 parechinus
23	90	43.1	180	5	025636	025636 parechinus
24	90	43.1	483	5	096148	096148 plasmodium
25	90	43.1	529	4	09NPX4	09NPX4 homo sapien
26	90	43.1	628	4	09H5M5	09H5M5 homo sapien
27	90	43.1	749	5	0967D9	0967D9 drosophila
28	90	43.1	1062	5	0960C4	0960C4 drosophila
29	90	43.1	1430	5	09W0B0	09W0B0 drosophila
30	90	43.1	1616	4	096PH3	096PH3 homo sapien
31	89	42.6	83	11	09ER82	09ER82 mus musculus
32	89	42.6	158	10	09MBF7	09MBF7 illium long
33	89	42.6	208	5	093901	093901 caenorhabdi
34	89	42.6	265	10	039598	039598 chlamydomon
35	88	42.1	260	11	0991X5	0991X5 mus musculus
36	88	42.1	392	5	018350	018350 caenorhabdi
37	88	42.1	409	5	090356	090356 caenorhabdi
38	88	42.1	1166	5	09V114	09V114 drosophila
39	87.5	41.9	533	10	09MAB3	09MAB3 arabidopsis
40	87	41.6	107	10	09LOF6	09LOF6 arabidopsis
41	87	41.6	565	10	09LD90	09LD90 arabidopsis
42	86	41.1	118	13	091A73	091A73 paralichthy
43	86	41.1	347	2	09RKL9	09RKL9 streptomyces
44	86	41.1	455	5	061747	061747 caenorhabdi
45	85.5	40.9	1173	5	095XK7	095XK7 caenorhabdi

#### ALIGNMENTS

RESULT	1	ALIGNMENTS
ID	09VG05	PRELIMINARY; PRT; 2451 AA.
AC	09VG05;	
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE	CG7518 PROTEIN.	
OS	CG7518.	
OC	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-BERKELEY;	
RX	MEDLINE=20196006; PubMed=10711132;	
RA	Adams M.D., Ceolniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,	
RA	Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dierck S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,	
RA	Glodes K., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,	
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigbam C.,	
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	







Query Match 47.8%; Score 100; DB 3; Length 128;  
Best Local Similarity 100.0%; Pred. No. 0.035;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 KKKKKKKKKKKKKKKKK 45  
DB 71 KKKKKKKKKKKKKKKKK 90

## RESULT 9

ID Q9H5V6 PRELIMINARY; PRT; 168 AA.  
AC Q9H5V6;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE CDNA: FLJ22976 FIS, CLONE KAT11222 (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
RA Suzuki Y., Ohtsubashi M., Nishi T., Shibahara T., Tanaka T.,  
RA Nakamura Y., Isogai T., Sugano S.;  
RT "NEDO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AK026629; BAB1513.1; -;  
FT NON\_TER 168  
SQ SEQUENCE 168 AA; 19549 MW; A19DBD195F8A1A90 CRC64;

Query Match 47.8%; Score 100; DB 4; Length 168;  
Best Local Similarity 100.0%; Pred. No. 0.044;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 KKKKKKKKKKKKKKKKK 45  
DB 140 KKKKKKKKKKKKKKKKK 159

## RESULT 10

ID Q64075 PRELIMINARY; PRT; 215 AA.  
AC Q64075;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE NUCLEOPORIN P62 HOMOLOG PROTEIN (FRAGMENT).  
OS Rattus sp.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10118;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95151924; PubMed=7849178;  
RA Wang Z.O., Akmal K.M., Kim K.H.;  
RT "An unusual nucleoporin-related messenger ribonucleic acid is present  
in the germ cells of rat testis."  
RL Biol. Reprod. 51:1022-1030(1994).  
DR EMBL: S75997; AAB3384.1; -;  
KW Porin.  
FT NON\_TER 1  
SQ SEQUENCE 215 AA; 24593 MW; 098251C97A8FBD8 CRC64;

Query Match 47.8%; Score 100; DB 11; Length 215;  
Best Local Similarity 100.0%; Pred. No. 0.055;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 KKKKKKKKKKKKKKKKK 45  
DB 35 KKKKKKKKKKKKKKKKK 54

## RESULT 11

ID Q9LG29 PRELIMINARY; PRT; 260 AA.  
AC Q9LG29;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE:PID9.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-COLUMBIA;  
RA Nakamura Y.;  
RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III."  
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AP002460; BAA97098.1; -;  
DR InterPro: IPR001386; Linker\_histone.  
DR PRINTS: PRO0624; HISTONEH5.  
SQ SEQUENCE 260 AA; 33307 MW; 43E2394CB8131143 CRC64;

Query Match 47.8%; Score 100; DB 10; Length 260;  
Best Local Similarity 100.0%; Pred. No. 0.066;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 KKKKKKKKKKKKKKKKK 45  
DB 7 KKKKKKKKKKKKKKKKK 26

## RESULT 12

ID Q9NT34 PRELIMINARY; PRT; 380 AA.  
AC Q9NT34;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE HYPOTHETICAL 42.7 KDA PROTEIN (FRAGMENT).  
GN DKFZP434I1120.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA Oltensweiler B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AL137556; CAB70810.1; -;  
KW Hypothetical protein.  
FT NON\_TER 380  
SQ SEQUENCE 380 AA; 42689 MW; 67F5DD101346AFB CRC64;

Query Match 47.8%; Score 100; DB 4; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.092;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 KKKKKKKKKKKKKKKKK 45  
DB 355 KKKKKKKKKKKKKKKKK 374

## RESULT 13

ID	ORGANISM	PRELIMINARY:	PRT:	377 AA.
AC	09GN12			
DT	01-MAR-2001	(TREMBlrel. 16, Created)		
DT	01-MAR-2001	(TREMBlrel. 16, Last sequence update)		
DE	01-DEC-2001	(TREMBlrel. 19, Last annotation update)		
GN	L3836.2.			
OS	Leishmania major.			
OC	Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.			
OX	NCBI_TaxID=5664;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-FRIEDLIN;			
RA	Myler P.J., Sisk E., Cawthra J., Handley F., Vogt C., Robertson L.,			
RA	McDonagh P., Stuart K., Worthey E.A.;			
RL	Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AC018847; AAG29141.1; -			
DR	InterPro: IPR001014; Antifreeze_1.			
DR	InterPro: IPR001304; lectin_C.			
DR	InterPro: IPR000306; znf_FYVE.			
DR	Pfam: PF01363; FYVE; 1.			
DR	PRINTS: PRO0308; ANTI-FREEZE1.			
DR	SMART: SM00064; FYVE; 1			
DR	PROSITE: PS00615; C_Type_Lectin_1; UNKNOWN_1.			
SO	SEQUENCE 377 AA; 43414 MW; 180C2F1DE47C5D7 CRC64;			
Query Match				
Best Local Similarity		47.6%;	Score 99.5;	DB 5; Length 377;
Matches 23; Conservative 13; Mismatches 6; Indels 5; Gaps 1		48.9%;	Pred. No. 0.1;	
OY	3	AAAAAEEAAAEAAAAA-----EAAAAAKKKKKKKKKKKKKKKKK 44		
DB	150	AAAAAEEAAAEAAAAAEEAAAEAAAAAEEERORAKKKRRRREAR 196		
RESULT 14				
Q9NKF0		PRELIMINARY:	PRT:	686 AA.
AC	Q9NKF0.			
DT	01-OCT-2000	(TREMBlrel. 15, Created)		
DT	01-OCT-2000	(TREMBlrel. 15, Last sequence update)		
DE	01-DEC-2001	(TREMBlrel. 19, Last annotation update)		
DE	CMDA.FLJ20288.FIS.CLONE.HEP04414 (FRAGMENT).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RA	Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,			
RA	Nakashima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,			
RA	Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isigai T., Sugano S.;			
RT	"NDDO human cDNA sequencing project."			
RL	Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AK000295; BAA01063.1; -			
DR	HSSP: 000420; IAWC			
DR	InterPro: IPR002110; ANK.			
DR	Pfam: PFO0023; ank; 14.			
DR	SMART: SM00248; ANK; 14.			
DR	PROSITE: PSS0088; ANK_REPEAT; 13.			
DR	PROSITE: PSS0297; ANK_REPEAT_REGION; 1.			
KW	ANK repeat; Repeat.			
FT	NON_TER 686			
SO	SEQUENCE 686 AA; 73615 MW; AE1380212E0FAD22 CRC64;			

[illegible]

Search completed: July 1, 2002, 06:30:59  
Job time: 691 sec

Query Match	47.4%	Score 99;	DB 4;	Length 686;
Best Local Similarity	65.6%;	Pred. No. 0.19;		
Matches 21; Conservative	4;	Mismatches	7;	Indels 0;
Gaps	0;			



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:24 ; Search time 98.97 Seconds  
(without alignments)  
50.503 Million cell updates/sec

Title: US-09-461-684-4

Sequence: 1 CEAAAAAAAAAAAAAAAAE.....KKKKKKKKKKKKKKKKKK 45

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008  
Listing first 45 summaries

## Database :

A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	209	100.0	45	21	AA13783	Soluble tandem p6A
2	119	56.9	36	21	AA13785	Ask21 linker pepit
3	119	56.9	630	21	AA13781	Modified fibre pro
4	119	56.9	640	21	AA13783	Modified fibre pro
5	118	56.5	2451	22	AA13783	Drosophila melanog
6	115.5	55.3	123	22	AA13783	Human polypeptide
7	115	55.0	113	22	AA13783	Human polypeptide
8	111	53.1	63	22	AA13783	Novel human DNA-bi
9	110	52.6	141	22	AA13783	Novel human DNA-bi
10	109	52.2	25	21	AA13781	Soluble peptide an
11	109	52.2	59	21	AA13781	Nuclear ligand use

12	109	52.2	59	21	AA13783	Nuclear ligand for
13	109	52.2	59	22	AA13783	Nucleic acid trans
14	109	52.2	59	22	AA13783	Nucleic acid trans
15	109	52.2	74	22	AA13783	Human polypeptide
16	109	52.2	75	22	AA13783	Human polypeptide
17	109	52.2	112	22	AA13783	Human polypeptide
18	109	52.2	150	22	AA13783	Human polypeptide
19	108	51.7	80	22	AA13783	Human polypeptide
20	108	51.7	126	22	AA13783	Human polypeptide
21	107	51.2	48	22	AA13783	Human polypeptide
22	107	51.2	63	22	AA13783	Human polypeptide
23	107	51.2	83	22	AA13783	Human polypeptide
24	106	50.7	26	22	AA13783	Human polypeptide
25	106	50.7	57	22	AA13783	Human polypeptide
26	106	50.7	64	22	AA13783	Human polypeptide
27	106	50.7	74	22	AA13783	Human polypeptide
28	106	50.7	74	22	AA13783	Human polypeptide
29	106	50.7	128	21	AA13783	Novel human DNA-bi
30	106	50.7	272	21	AA13783	Novel human DNA-bi
31	105	50.2	27	21	AA13783	Novel human DNA-bi
32	105	50.2	31	22	AA13783	Novel human DNA-bi
33	105	50.2	62	20	AA13783	Novel human DNA-bi
34	105	50.2	62	22	AA13783	Novel human DNA-bi
35	105	50.2	66	22	AA13783	Novel human DNA-bi
36	105	50.2	66	22	AA13783	Novel human DNA-bi
37	105	50.2	72	22	AA13783	Novel human DNA-bi
38	105	50.2	83	22	AA13783	Novel human DNA-bi
39	105	50.2	84	22	AA13783	Novel human DNA-bi
40	105	50.2	122	22	AA13783	Novel human DNA-bi
41	105	50.2	155	22	AA13783	Novel human DNA-bi
42	105	50.2	175	11	AA13783	Novel human DNA-bi
43	105	50.2	630	21	AA13783	Novel human DNA-bi
44	105	50.2	640	21	AA13783	Novel human DNA-bi
45	104	49.8	28	22	AA13783	Novel human DNA-bi

## ALIGNMENTS

RESULT 1  
AA13783  
ID AA13783 standard; peptide: 45 AA.  
XX  
AC AA13783:  
X  
DT 10-NOV-2000 (first entry)  
XX  
DE Soluble tandem p6A/ p6 peptide conjugate.  
XX  
KW pk peptide: cytosolic; vaccine; cytotoxic T cell; CTL; immunotherapy;  
KW major histocompatibility complex class I; MHC class I; antigen; tumour;  
KW prostate; breast; multiple myeloma; p6A peptide.  
XX  
OS Unidentified.  
XX  
PN WO200035949-A1.  
XX  
PD 22-JUN-2000.  
XX  
PF 14-DEC-1999: 99WO-US29724.  
XX  
PR 14-DEC-1998: 98US-0112324.  
XX  
(DEND-) DENDREON CORP.  
XX  
PI Laus R, Hakim I, Vidovic D;  
XX  
DR WPI, 2000-442365/38.  
XX  
PT Antigens modified by the covalent addition of a peptide that  
PT facilitates entry into antigen presenting cells, useful for producing  
PT compositions for immunizing against tumors and pathogens -

PS Claim 2; Page 26; 34pp; English.

xx The present invention relates to compositions of modified soluble protein  
CC antigens capable of eliciting an enhanced *in vivo* cytotoxic T cell (CTL)  
CC response i.e. a major histocompatibility complex (MHC) class I molecule  
CC response. The protein antigen is modified by the covalent addition of a  
CC peptide sequence which facilitate entry of the antigen into antigen  
CC presenting cells (APCs). The present sequence is one such peptide  
CC sequence which can be used to modify the soluble antigens. The present  
CC sequence is tandem pEA/ PK peptide conjugate. The modified antigen  
CC composition may be used for immunising against, or treating a tumour e.g.  
CC prostate and breast carcinoma or multiple myeloma, or pathogen in  
CC mammals.

SO Sequence 45 AA;

Qy 1 CEAA 45  
Db 1 ccaaaaaaaaaaaaaaaaaaaaaaaaaaakkkkkkkkkkkkkkkkkkkkkkkkkkkkk 45

RESULT 2  
AAB23585  
ID AAB23585 standard; peptide: 36 AA.

AC AAB23585;  
XX  
XX 11-JAN-2001 (first entry)  
DT  
XX  
XX Ask21 linker peptide.  
DE  
XX  
XX Viral vector: melanocyte-stimulating hormone receptor; MSH; cytostatic;  
KW tumour; malignant melanoma; linker.  
KM  
OS Synthetic.  
XX  
XX WO200050618-A1.  
PN  
XX 31-AUG-2000.  
PD  
XX 24-FEB-2000; 2000WO-JP01069.  
PE  
XX 24-FEB-1999; 99JP-0093263.  
PR  
XX (NICA-) JAPANESE FOUND CANCER RES.  
PA  
XX Hamada H;  
PI  
XX WPI: 2000-549414/50.  
DR  
XX N-PSDB: AAA93817.

PT Virus vector useful in diagnosis and treatment of tumors particularly  
PT malignant melanoma, constructed by fusing viral protein with ligand  
PT binding specifically to melanocyte-stimulating hormone receptor  
XX  
XX  
XX Claim 4; Page 100; 145pp; Japanese.

CC This invention relates to a viral vector constructed by the fusion of a  
CC viral protein with a ligand which binding specifically to the  
CC melanocyte-stimulating hormone (MSH) receptor. The vector contains one  
CC of four linkers represented by sequences AAA93815-A93818 and  
CC AAB23583-823586, and DNA encoding a fibre protein selected from those  
CC represented by AAA93819-A93826 and AAB23587-823594. The vector has  
CC cytostatic activity, and can be used for gene therapy and in the  
CC diagnosis and treatment of tumours, particularly malignant melanomas. The  
CC present sequence represents an Ask21 linker DNA sequence used in the  
CC construction of the vector.

```

SQ      Sequence      36 AA:

Query Match          56.9%: Score 119; DB 21; Length 36;
Best Local Similarity 70.6%: Pred. No. 1e-05;
Matches 24; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY      12 AAEAAAAAEEEEAAKKKKKKKKKKKKKKKKKKKKK 45
       :|::||: : |||||||
Db       2 sasasasapgsqskkkkkkkkkkkkkkkkkkkk 35

RESULT      3
AAB23591
ID   AAB23591 standard; Protein; 630 AA.
XX
AC   AAB23591;
XX
DT   11-JAN-2001 (first entry)
DE   Modified fibre protein encoded in pWE6.7R-F/ask2JMSHa.
XX
KW   Viral vector; melanocyte-stimulating hormone receptor; MSH; cytostatic;
KM   tumour; malignant melanoma; fibre protein.
XX
OS   Synthetic.
PN   WO200050618-A1.
XX
PD   31-AUG-2000.
XX
PE   24-FEB-2000; 2000WO-JP01069.
PR   24-FEB-1999; 99JP-0093263.
XX
PA   (NICA-) JAPANESE FOUND CANCER RES.
XX
PI   Hamada H;
DR   WPI; 2000-549414/50.
DR   N-PSDB; AAA93823.
XX

Virus vector useful in diagnosis and treatment of tumors particularly
PT   malignant melanoma, constructed by fusing viral protein with ligand
PR   binding specifically to melanocyte-stimulating hormone receptor -
XX
PS   Claim 24; Page 121-126; 145pp; Japanese.
XX

This invention relates to a viral vector constructed by the fusion of a
CC   viral protein with a ligand which binds specifically to the
CC   melanocyte-stimulating hormone (MSH) receptor. The vector contains one
CC   of four linkers represented by sequences AAA93815-A93818 and
CC   AAA23383-B23386, and DNA encoding a fibre protein selected from those
CC   represented by AAA93819-A93826 and AAB23587-B23594. The vector has
CC   cytosolic activity, and can be used for gene therapy and in the
CC   diagnosis and treatment of tumours, particularly malignant melanomas. The
CC   present sequence represents a modified fibre protein used in the
CC   construction of the vector.
XX
SQ      Sequence      630 AA:

Query Match          56.9%: Score 119; DB 21; Length 630;
Best Local Similarity 70.6%: Pred. No. 0.00016;
Matches 24; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY      12 AAEAAAAAEEEEAAKKKKKKKKKKKKKKKKKKKKK 45
       :|::||: : |||||||
Db       583 sasasasapgsqskkkkkkkkkkkkkkkkkkkk 616

RESULT      4
AAB23593
```

ID AAB23593 standard; Protein: 640 AA.  
XX AAB23593;  
AC  
XX  
XX 11-JAN-2001 (first entry)  
DT  
XX  
XX Modified fibre protein encoded in pME6.7R-F/ask21MSH.  
DE  
XX  
XX Viral vector: melanocyte-stimulating hormone receptor; MSH; cytostatic;  
KW tumour; malignant melanoma; fibre protein.  
XX  
XX Synthetic.  
OS  
XX  
XX WO200050618-A1.  
PN  
XX  
XX 31-AUG-2000.  
PD  
XX  
XX 24-FEB-2000; 2000WO-JP01069.  
PF  
XX  
XX 24-FEB-1999; 99JP-0093263.  
PR  
XX  
XX (NICA-) JAPANESE FOUND CANCER RES.  
PA  
XX  
XX Hamada H;  
PI  
XX  
XX WPI: 2000-549414/50.  
DR  
XX  
XX N-PSDB; AAB235825.  
NR  
XX  
XX Virus vector useful in diagnosis and treatment of tumors particularly  
PT malignant melanoma, constructed by fusing viral protein with ligand  
PT binding specifically to melanocyte-stimulating hormone receptor  
XX  
XX Claim 24; Page 131-136; 145pp; Japanese.  
PS  
XX  
XX This invention relates to a viral vector constructed by the fusion of a  
CC viral protein with a ligand which binding specifically to the  
CC melanocyte-stimulating hormone (MSH) receptor. The vector contains one  
CC of four linkers represented by sequences AAB235815-A93818 and  
CC AAB23583-B23586, and DNA encoding a fibre protein selected from those  
CC represented by AAB23587-A93826 and AAB23587-B23594. The vector has  
CC cytostatic activity, and can be used for gene therapy and in the  
CC diagnosis and treatment of tumours, particularly malignant melanomas. The  
CC present sequence represents a modified fibre protein used in the  
CC construction of the vector.  
XX  
XX Sequence 640 AA:  
SQ

Query Match 56.9%; Score 119; DB 21; Length 640;  
Best Local Similarity 70.6%; Pred. No. 0.00016;  
Matches 24; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 12 AAEAAAAAEAAA 45  
DB 583 sasasasapsagsakkkkkkkkkkkkkkkkkkkkkkkkkkkkk 616

RESULT 5  
ID ABB71574  
AB 71574 standard; Protein: 2451 AA.  
AC  
XX  
XX ABB71574;  
AC  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 41514.  
DE  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX WO200171042-A2.  
PN

XX 27-SEP-2001.  
PD  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX  
XX 11-JUL-2000; 2000US-0614150.  
PR  
XX  
XX (PEKE ) PE CORP NY.  
PA  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX  
XX WPI: 2001-656860/75.  
DR  
XX  
XX N-PSDB; ABL15677.  
NR  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX  
XX Disclosure; SEQ ID NO 41514; 21pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL10511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 2451 AA:  
SQ

Query Match 56.5%; Score 118; DB 22; Length 2451;  
Best Local Similarity 75.7%; Pred. No. 0.00072;  
Matches 28; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 3 AAEAAAAAEAAA 39  
DB 1374 aaaaaaaaaaaaaaaaaaaaaaaaaaegaklkknkkgakk 1410

RESULT 6  
ID AAO01368  
AA 01368 standard; Protein: 123 AA.  
AC  
XX  
XX AAO01368;  
AC  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX  
XX Human polypeptide SEQ ID NO 15260.  
DE  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200164835-A2.  
PN  
XX  
XX 07-SEP-2001.  
PD  
XX  
XX 26-FEB-2001; 2001WO-US04927.  
PF  
XX  
XX 28-FEB-2000; 2000US-0515126.  
PR  
XX  
XX 18-MAY-2000; 2000US-0577409.  
PR  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX

P1	Tang YT, Liu C, Drmanac RT;
XX	
DR	WPI: 2001-514838/56.
DR	N-PSDB: AAI81299.
XX	
PT	Isolated nucleic acids and polypeptides, useful for preventing
PT	diagnosing and treating e.g. leukaemia, inflammation and immune
PT	disorders -
PS	Claim 20; SEQ ID NO 15260; 1399pp + Sequence listing; English.
XX	
CC	The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC	the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC	cytokine, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cytokine-like activities,
CC	e.g. stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	treatment/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	inflammation.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
CC	
XX	
SQ	Sequence 123 AA;
	Query Match 55.3%; Score 115.5; DB 22; Length 123;
	Best Local Similarity 58.7%; Pred. No. 7.3e-05;
	Matches 27; Conservative 5; Mismatches 11; Indels 3; Gaps 1;
Oy	3 AAAAEEAAAEEAAAEEAAA--AAKKKKKKKKKKKKKKKKK 45
	: : :   : : :   : : :   : : :   : : :   : : :
Dd	61 apseagsigaadsgpaatlparqiltkttkkkkkkkkkkkkkkk 106
RESULT 7	
AAO03703	
ID	AAO03703 standard; Protein; 113 AA.
XX	
AC	AAO03703:
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human polypeptide seq ID NO 17595.
XX	
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW	vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;
KW	nervous system disorders; arthritis; inflammation.
XX	
OS	Homo sapiens.
XX	
PN	.MO200164835-A2.
PD	
XX	07-SEP-2001.
XX	
PF	26-FEB-2001; 2001WO-USO4927.
XX	
PR	28-FEB-2000; 2000US-0515126.
XX	
PR	18-MAY-2000; 2000US-0577409.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
XX	Tang YT, Liu C, Drmanac RT;
PI	
XX	WPI: 2001-514838/56.
DR	I-PSDB: AAI83634.
DR	
XX	
PT	Isolated nucleic acids and polypeptides, useful for preventing
PT	diagnosing and treating e.g. leukaemia, inflammation and immune

```

PT disorders - -
XX
PS Clalm 20; SEQ ID NO 17595; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapies. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 113 AA:

Query Match 55.0%; Score 115; DB 22; Length 113;
Best Local Similarity 70.3%; Pred. No. 7.6e-05;
Matches 26; Conservative 0; Mismatches 11; Indels 0; Gaps 0

QY 9 AAAAAEAAAAAAAAAKKKKKKKKKKKKKKKKKK 45
   | | | | | | | | | | | | | | | | | | |
Db 46 ataagexiwdaekkkkkkkkkkkkkkkkkk 82

RESULT 8
AAU18275
ID AAU18275 standard; Protein; 63 AA.
XX
AC AAU18275;
XX
DT 21-NOV-2001 (first entry)
XX
DE Novel human DNA-binding protein #122.
XX
KW Human; DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; Y-box binding protein;
KW DNA organisation; gene transcription; malignant disease;
KW autoimmune disorder; rheumatic disease; genetic abnormality;
KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW cyclostatic.
XX
OS Homo sapiens.
XX
PN WO20015162-A1.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01305.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0214886.
PR 28-JUN-2000; 2000US-0215135.
PR 30-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.

```

PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234597.  
PR 25-SEP-2000; 2000US-0234598.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235634.  
PR 27-SEP-2000; 2000US-0235636.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.

PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-465557/50.  
N-PSDB; AAS29151.  
  
XX Nucleic acid molecules encoding human secreted chromosomal binding  
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
PT Alzheimer's and Parkinson's diseases and cancers -  
XX  
XX  
PS Claim 11; SEQ ID No 260; 561pp; English.  
XX  
CC The present invention relates to the isolation of novel DNA-binding  
CC proteins, and cDNA (AAS29030-AAS29157) and genomic sequences encoding  
CC for these proteins. DNA-binding proteins such as histones, chromo  
CC (chromatin organisation modifier) domain proteins, and Y-box binding  
CC proteins may contribute to diseases resulting from aberrant DNA  
CC organisation and/or gene transcription. The sequences of the invention  
CC are useful in screening assays to identify antagonists and/or agonists  
CC that may enhance or block activities mediated by DNA-binding proteins.  
CC Blockers of DNA-binding proteins may be useful in treating disorders  
CC such as malignant diseases (e.g. cancer), autoimmune disorders  
CC (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid  
CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious





## RESULT 12

AC AAU04285;

XX 23-OCT-2001 (first entry)  
XX  
DE Nuclear ligand #2 used in nucleic acid transporter system.  
XX  
KW Nucleic acid transport; cytosol; ligand; lysis agent; spacer molecule;  
KM gene therapy; hepatocyte; muscle; bone forming cell.  
XX  
OS Synthetic.  
XX  
PN US6177554-B1.  
XX  
PD 23-JAN-2001.  
XX  
PF 05-JUN-1995; 95US-0462040.  
XX  
PR 14-DEC-1993; 93US-0167641.  
PR 20-MAR-1992; 92US-0855389.  
PR 19-MAR-1993; 93WO-US02725.  
XX  
PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
PI Woo SLC, Smith LC, Cristiano RJ, Gottchalk S, Sparrow J;  
XX  
DR WPI; 2001-36593/38.  
XX  
PT Nucleic acid transport system, useful for creating transgenic animals  
XX for assessing human disease such as cancer in an animal model -  
XX  
PS Disclosure; Column 8; 11pp; English.  
XX  
CC The sequence represents the nuclear ligand, #2, used in a nucleic acid  
CC transporter system. The nucleic acid transporter system uses nucleic acid  
CC binding complexes containing surface ligands which are capable of binding  
CC to a cell surface receptor and entering the cell through cytosol. The  
CC compounds of the invention are either ligands, binding molecules (surface  
CC ligands), lysis agents, spacer molecules or their intermediates. The  
CC ligands, binding molecules, lysis agents and spacer molecules are used in  
CC nucleic acid transporter systems to deliver nucleic acid into specific  
CC cells e.g. in gene therapy to deliver nucleic acid into hepatocytes,  
CC muscle cells or bone forming cells.  
XX  
SO Sequence 59 AA;  
  
Query Match 52.2%; Score 109; DB 22; Length 59;  
Best Local Similarity 88.5%; Pred. No. 0.00016;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 20 EAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 45  
II |  
DB 12 eapyyakakakakakakakakakakak 37  
  
RESULT 14  
ID AAB45848 standard; Protein: 59 AA.  
XX  
AC AAB45848;  
XX  
DT 21-MAR-2001 (first entry)  
XX  
DE Nucleic acid transporter system peptide ligand SEQ ID NO 60.  
XX  
KW Nucleic acid delivery; nucleic acid transporter system; hormone; enzyme;  
KM growth factor; clotting factor; apolipoprotein; receptor; drug; oncogene;  
KM tumor antigen; tumor suppressor; viral antigen; parasitic antigen;  
XX  
OS unidentified.  
XX  
PN US6150168-A.  
XX

PD 21-NOV-2000.  
XX  
PF 05-JUN-1995; 95US-0460971.  
XX  
PR 14-DEC-1993; 93US-0167641.  
PR 20-MAR-1992; 92US-0855389.  
PR 19-MAR-1993; 93WO-US02725.  
XX  
PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
PI Gottchalk S, Sparrow J, Cristiano RJ, Smith LC, Woo SLC;  
XX  
DR WPI; 2001-049093/06.  
XX  
PT Nucleic acid transporter system for delivering nucleic acid into a  
XX cell, useful for delivering proteins and polypeptides to cells,  
XX including growth factors, enzymes, hormones, and tumor suppressors -  
XX  
PS Disclosure; Column 123-124; 105pp; English.  
XX  
CC This invention describes a novel system (1) for delivering a nucleic acid  
CC to a cell, comprising a binding complex comprising a ligand binding  
CC molecule noncovalently bound to a nucleic acid and covalently linked to a  
CC surface ligand, and a second binding complex comprising a second binding  
CC molecule noncovalently bound to a nucleic acid and covalently linked to a  
CC nuclear ligand. The complexes are simultaneously bound to the nucleic  
CC acid. The nucleic acid transporter system can also be used in a method  
CC for the in vivo targeting of the insertion of DNA into a cell. It can  
CC also be used in processes for producing transformed cell lines. The  
CC system can be used to deliver a variety of proteins and polypeptides,  
CC such as hormones, growth factors, enzymes, clotting factors,  
CC apolipoproteins, receptors, drugs, oncogenes, tumor antigens, tumor  
CC suppressors, viral antigens, parasitic antigens, and bacterial antigens.  
CC The transporter system uses lysis agents to overcome the problems of  
XX endosomal/lysosomal degradation seen with prior art systems.  
XX  
SO Sequence 59 AA;  
  
Query Match 52.2%; Score 109; DB 22; Length 59;  
Best Local Similarity 88.5%; Pred. No. 0.00016;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 20 EAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 45  
II |  
DB 12 eapyyakakakakakakakakakakak 37  
  
RESULT 15  
ID AAO03278 standard; Protein: 74 AA.  
XX  
AC AAO03278;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 17170.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 1, 2002, 06:28:12 ; Search time 49.45 Seconds

(without alignments)  
85.499 Million cell updates/sec

Title: US-09-461-684-5

Sequence: 1 GLEFGAIFAGFLENGEGMID.....KKKKKKKKKKKKKKKKKK 44

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	55.1	550	1 HMTVS2	hemagglutinin prec
2	134	55.1	550	1 HMTVS3	hemagglutinin prec
3	134	55.1	550	1 HMTV77	hemagglutinin prec
4	134	55.1	550	1 HMTV80	hemagglutinin prec
5	134	55.1	550	1 HMTV33	hemagglutinin prec
6	134	55.1	550	1 HMTV89	hemagglutinin prec
7	134	55.1	550	1 HMTV21	hemagglutinin prec
8	134	55.1	550	1 HMTV98	hemagglutinin prec
9	134	55.1	550	1 HMTV15	hemagglutinin prec
10	134	55.1	550	2 JQ1153	hemagglutinin prec
11	134	55.1	550	2 JQ1154	hemagglutinin prec
12	134	55.1	550	2 JQ1155	hemagglutinin prec
13	134	55.1	566	1 HMTVH	hemagglutinin prec
14	134	55.1	566	1 HMTVHA	hemagglutinin prec
15	134	55.1	566	1 HMTVHM	hemagglutinin prec
16	134	55.1	566	1 HMTVHU	hemagglutinin prec
17	134	55.1	570	1 A45591	hemagglutinin prec
18	134	55.1	570	2 S22013	hemagglutinin prec
19	134	54.7	561	1 HMTV49	hemagglutinin prec
20	133	54.7	561	1 HMTV84	hemagglutinin prec
21	132	54.3	565	1 HMTV81	hemagglutinin prec
22	132	54.3	565	1 HMTV83	hemagglutinin prec
23	132	54.3	566	1 HMTV8	hemagglutinin prec
24	132	54.3	566	1 HMTV6	hemagglutinin prec
25	131	53.9	562	2 S38637	hemagglutinin - In
26	131	53.9	550	1 HMTV86	hemagglutinin prec
27	131	53.9	560	1 HMTV7	hemagglutinin prec
28	131	53.9	565	1 HMTV2	hemagglutinin prec
29	131	53.9	565	1 HMTV4	hemagglutinin prec

30	131	53.9	565	1 HMTV5	hemagglutinin prec
31	131	53.9	565	1 HMTV6	hemagglutinin prec
32	131	53.9	565	1 HMTV7	hemagglutinin prec
33	131	53.9	565	1 HMTV8	hemagglutinin prec
34	131	53.9	565	1 HMTV9	hemagglutinin prec
35	131	53.9	565	1 HMTVET	hemagglutinin prec
36	131	53.9	565	1 HMTVEE	hemagglutinin prec
37	131	53.9	565	2 S33703	hemagglutinin - In
38	131	53.9	570	2 S22014	hemagglutinin prec
39	131	53.9	570	2 S22015	hemagglutinin prec
40	131	53.9	570	2 S22016	hemagglutinin prec
41	131	53.9	570	2 S22017	hemagglutinin prec
42	131	53.9	570	2 S22018	hemagglutinin prec
43	131	53.9	570	2 S22020	hemagglutinin prec
44	131	53.9	570	2 S22021	hemagglutinin prec
45	131	53.9	570	2 S22029	hemagglutinin prec

## ALIGNMENTS

```
RESULT 1
HMTVS2 hemagglutinin precursor - Influenza A virus (strain A/swine/126/82) (fragment)
C:Species: Influenza A virus
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C:Accession: A29971
R:Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A:Title: Origin of the hemagglutinin gene of H3N2 Influenza viruses from pigs in China
A:Reference number: A94370; M0ID:88101364
A:Accession: A29971
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M19056; NID:g324208
A>Note: the sequence in GenBank entry FLAHAP, release 106, (PID:g324209) differs from
C:Genetics:
A:Map position: segment 4
C:Superfamily: Influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:130-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1% Score 134: DB 1: Length 550:
Best Local Similarity 100.0%: Pred. No. 3.1e-06:
Matches 23: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLEFGAIFAGFLENGEGMIDGWC 24
DB 330 GLEFGAIFAGFLENGEGMIDGWC 352

RESULT 2
HMTVS3 hemagglutinin precursor - Influenza A virus (strain A/swine/81/78) (fragment)
C:Species: Influenza A virus
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C:Accession: B29971
R:Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A:Title: Origin of the hemagglutinin gene of H3N2 Influenza viruses from pigs in China
A:Reference number: A94370; M0ID:88101364
A:Accession: B29971
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M19057; NID:g324210
A>Note: the sequence in GenBank entry FLAHAP, release 106, (PID:g324211) differs from
C:Genetics:
```

A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-338/Product: hemagglutinin HA1 #status predicted <HA1>  
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1% Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFTENGMEGMDGWC 24  
DB 330 GLFGAAGFTENGMEGMDGWC 352

RESULT 3  
HMIv77  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: A27813  
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: A27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16737; NID:9324081; PIDN:AAA43143.1; PID:9324082  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-338/Product: hemagglutinin HA1 #status predicted <HA1>  
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1% Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFTENGMEGMDGWC 24  
DB 330 GLFGAAGFTENGMEGMDGWC 352

RESULT 4  
HMIv80  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: B27813  
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: B27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16738; NID:9324083  
A:Note: The translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1  
C:Genetics:

A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-338/Product: hemagglutinin HA1 #status predicted <HA1>  
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1% Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFTENGMEGMDGWC 24  
DB 330 GLFGAAGFTENGMEGMDGWC 352

RESULT 5  
HMIv33  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: C27813  
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: C27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16739; NID:9324085; PIDN:AAA43145.1; PID:9324086  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-338/Product: hemagglutinin HA1 #status predicted <HA1>  
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1% Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFTENGMEGMDGWC 24  
DB 330 GLFGAAGFTENGMEGMDGWC 352

RESULT 6  
HMIv89  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: D27813  
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: D27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16740; NID:9324087; PIDN:AAA43146.1; PID:9324088  
C:Genetics:  
A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HAI #status predicted <HA1>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLEFAGIAGFIENGMEGMDGWC 24  
DB 330 GLEFAGIAGFIENGMEGMDGWC 352

RESULT 7  
HMIY21  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)  
N:Contains: hemagglutinin HAI; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: E27813  
R:Kida, H.; Kawacka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: E27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16741; NID:9324089  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer  
F:1-328/Product: hemagglutinin HAI #status predicted <HA1>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLEFAGIAGFIENGMEGMDGWC 24  
DB 330 GLEFAGIAGFIENGMEGMDGWC 352

RESULT 8  
HMIY98  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)  
N:Contains: hemagglutinin HAI; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: F27813  
R:Kida, H.; Kawacka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: F27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16742; NID:9324091  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HAI #status predicted <HA1>  
F:530-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLEFAGIAGFIENGMEGMDGWC 24  
DB 330 GLEFAGIAGFIENGMEGMDGWC 352

RESULT 9  
HMIY15  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)  
N:Contains: hemagglutinin HAI; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: G27813  
R:Kida, H.; Kawacka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: G27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16743; NID:9324093; PIDN:AAA3149.1; PID:9324094  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HAI #status predicted <HA1>  
F:520-536/Domain: transmembrane #status predicted <HA2>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLEFAGIAGFIENGMEGMDGWC 24  
DB 330 GLEFAGIAGFIENGMEGMDGWC 352

RESULT 10  
J01153  
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)  
N:Contains: hemagglutinin HAI; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000  
C:Accession: J01153  
R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.  
J. Gen. Virol. 72, 2007-2010, 1991  
A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian  
A:Reference number: J01153; MUID:91341491  
A:Accession: J01153  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <YAS>  
A:Cross-references: GB:D00929; NID:9221279; PIDN:BAA00769.1; PID:9221280  
A:Note: the authors translated the codon GCG for residue 218 as Glu  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; homotrimer  
F:1-328/Product: hemagglutinin HAI #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3,1e-06;  
Matches 23: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAAGFIENGEGMIDGWC 24  
|||||

Db 330 GLFGAAGFIENGEGMIDGWC 352

## RESULT 11

hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragment)  
N:Contains: hemagglutinin HA1: hemagglutinin HA2

C:Species: Influenza A virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000  
C:Accession: J01154

R:Yasuda, J.; Shortridge, K.F.; Shmizu, Y.; Kida, H.  
J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3  
A:Reference number: J01153; MUID:91341491

A:Accession: J01154

A:Molecule type: genomic RNA

A:Residues: 1-550 <YAS>

A:Cross-references: GB:D00930; NID:9221273; PIDN:BAA00770.1; PID:9221274

A:Note: the authors translated the codon GCG for residue 218 as Glu

A:Note: residues 528-532 are not shown in this publication

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: glycoprotein; homotrimer

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3,1e-06;

Matches 23: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAAGFIENGEGMIDGWC 24  
|||||

Db 330 GLFGAAGFIENGEGMIDGWC 352

## RESULT 12

hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment)  
N:Contains: hemagglutinin HA1: hemagglutinin HA2

C:Species: Influenza A virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000  
C:Accession: J01155

R:Yasuda, J.; Shortridge, K.F.; Shmizu, Y.; Kida, H.  
J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3  
A:Reference number: J01153; MUID:91341491

A:Accession: J01155

A:Molecule type: genomic RNA

A:Residues: 1-550 <YAS>

A:Cross-references: GB:D00931; NID:9221277; PIDN:BAA00771.1; PID:9221278

A:Note: the authors translated the codon GCG for residue 218 as Glu, GCC for residue 538

A:Note: residues 528-532 are not shown in this publication

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: glycoprotein; homotrimer

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3,1e-06;  
Matches 23: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAAGFIENGEGMIDGWC 24  
|||||

Db 330 GLFGAAGFIENGEGMIDGWC 352

## RESULT 13

HMIIVH

hemagglutinin precursor - influenza A virus  
C:Species: Influenza A virus

C:Date: 28-Feb-1981 #sequence\_revision 28-Feb-1981 #text\_change 22-Oct-1999  
C:Accession: A93703; A93233; A04051; A93231; A94441

R:Both, G.W.; Sleight, M.J.  
Nucleic Acids Res. 8, 2561-2575, 1980

A:Title: Complete nucleotide sequence of the haemagglutinin gene from a human influe  
A:Reference number: A93705; MUID:81053698

A:Accession: A93705

A:Molecule type: genomic RNA

A:Residues: 1-566 <BOT>

A:Cross-references: GB:V01103

A:Experimental source: strain A/NT/60/68/29C

A:Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/  
R:Doipheld, T.A.; Ward, C.W.

FEBS Lett. 110, 181-183, 1980

A:Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.  
A:Reference number: A91276; MUID:80179105

A:Contents: annotation; disulfide bonds

R:Gething, M.J.; Byr, J.; Skehel, J.J.; Waterfield, M.

Nature 287, 301-306, 1980

A:Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes  
A:Reference number: A93233; MUID:81030852

A:Accession: A93233

A:Molecule type: genomic RNA

A:Residues: 1-24, 'S', '26', 'D', '28-159', 'G', '161-197', 'I', '199-241', 'L', '243-249' <GET>

A:Experimental source: strain X-31[H3]

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>

F:536-552/Domain: transmembrane #status predicted <TM1>

F:30-482,68-293,80-92,155-489,297-321/Disulfide Bonds: #status experimental

F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 3,2e-06;  
Matches 23: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAAGFIENGEGMIDGWC 24  
|||||

Db 346 GLFGAAGFIENGEGMIDGWC 368

## RESULT 14

HMIIVH

hemagglutinin precursor - influenza A virus (strain A/Alchi/2/68)  
N:Contains: hemagglutinin HA1: hemagglutinin HA2

C:Species: Influenza A virus  
C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 16-Jul-1999  
C:Accession: A93231; A04051

R:Verhoeven, M.; Pang, R.; Min Jou, W.; Devos, R.; Huybreckx, D.; Saman, E.; Fiers  
Nature 286, 771-776, 1980

A:Title: Antigenic drift between the hemagglutinin of the Hong Kong influenza strai  
A:Reference number: A93231; MUID:80254693

A:Accession: A93231

A:Molecule type: genomic RNA

A:Residues: 1-566 <VER>

A:Cross-references: GB:J02090; NID:9324131; PIDN:AAA43178.1; PID:9324132

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>



F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>  
F:555,562/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 3.2e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAAGFIENGEGMIDGMYG 24  
|||||  
DB 346 GLFGAIAAGFIENGEGMIDGMYG 368

RESULT 15

HMIVHM  
hemagglutinin precursor - influenza A virus (strain A/Mem/102/72)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 31-Mar-2000  
C:Accession: A94441; A04051  
R:Seigh, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.  
In Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, B  
A:Title: The haemagglutinin gene of Influenza A virus: nucleotide sequence analysis of c  
A:Reference number: A94441  
A:Accession: A94441  
A:Molecule type: genomic RNA  
A:Residues: 1-566 <SLE>  
C:Superfamily: Influenza virus hemagglutinin  
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-16/Domain: signal sequence #status predicted <Sig>  
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>  
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>  
F:555,562/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 3.2e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAAGFIENGEGMIDGMYG 24  
|||||  
DB 346 GLFGAIAAGFIENGEGMIDGMYG 368

Search completed: July 1, 2002, 06:28:12  
Job time: 714 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 1, 2002, 06:31:35 ; Search time 25.37 Seconds

(Without alignments)  
67.153 Million cell updates/sec

Title: US-09-461-684-5

Sequence: 1 CGLFCAIAGFIENGWGMID.....KKKKKKKKKKKKKKKKKK 44

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	55.1	550	1	HEMA_IADH1
2	134	55.1	550	1	HEMA_IADH2
3	134	55.1	550	1	HEMA_IADH3
4	134	55.1	550	1	HEMA_IADH4
5	134	55.1	550	1	HEMA_IADH5
6	134	55.1	550	1	HEMA_IADH6
7	134	55.1	550	1	HEMA_IADH7
8	134	55.1	550	1	HEMA_IADH8
9	134	55.1	550	1	HEMA_IADH9
10	134	55.1	550	1	HEMA_IADH10
11	134	55.1	550	1	HEMA_IADH11
12	134	55.1	550	1	HEMA_IADH12
13	134	55.1	550	1	HEMA_IADH13
14	134	55.1	550	1	HEMA_IADH14
15	134	55.1	550	1	HEMA_IADH15
16	134	55.1	550	1	HEMA_IADH16
17	134	55.1	550	1	HEMA_IADH17
18	134	55.1	550	1	HEMA_IADH18
19	134	55.1	550	1	HEMA_IADH19
20	134	55.1	550	1	HEMA_IADH20
21	134	55.1	550	1	HEMA_IADH21
22	134	55.1	550	1	HEMA_IADH22
23	134	55.1	550	1	HEMA_IADH23
24	134	55.1	550	1	HEMA_IADH24
25	134	55.1	550	1	HEMA_IADH25
26	134	55.1	550	1	HEMA_IADH26
27	134	55.1	550	1	HEMA_IADH27
28	134	55.1	550	1	HEMA_IADH28
29	134	55.1	550	1	HEMA_IADH29
30	134	55.1	550	1	HEMA_IADH30
31	134	55.1	550	1	HEMA_IADH31
32	134	55.1	550	1	HEMA_IADH32
33	134	55.1	550	1	HEMA_IADH33

34	131	53.9	565	1	HEMA_IADH7	P16996 Influenza a
35	131	53.9	565	1	HEMA_IADH1	P15658 Influenza a
36	131	53.9	565	1	HEMA_IADH1	P16997 Influenza a
37	131	53.9	565	1	HEMA_IADH1	P16998 Influenza a
38	131	53.9	565	1	HEMA_IADH1	P16999 Influenza a
39	131	53.9	565	1	HEMA_IADH1	P17000 Influenza a
40	131	53.9	565	1	HEMA_IADH1	P17001 Influenza a
41	131	53.9	565	1	HEMA_IADH1	P17002 Influenza a
42	131	53.9	565	1	HEMA_IADH1	P17003 Influenza a
43	131	53.9	565	1	HEMA_IADH1	P17004 Influenza a
44	131	53.9	565	1	HEMA_IADH1	P17005 Influenza a
45	131	53.9	565	1	HEMA_IADH1	P17006 Influenza a

## ALIGNMENTS

RESULT 1  
HEMA\_IADH1 STANDARD: PRT: 550 AA.  
AC P12582: 084021; 084022;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hokkaido/5/77).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=11357;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87265458; PubMed=2440178;  
RA Kida H., Kawoka Y., Naeve C.W., Webster R.G.;  
RT "Antigenic and genetic conservation of H3 influenza virus in wild  
RT ducks";  
RL Virology 159:109-119(1987).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M16737; AAA43143.1; -  
CC PIR: A27813; HMTV77.  
CC InterPro: IPR001364; Hemagglutn.  
CC Pfam: PF00509; Hemagglutinin; 1.  
CC ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON TER 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA: 61705 MW: 767ACFE716FC969A CRC64;  
Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;

Matches 23: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAAGFIENGMEGMDGMYG 24  
|||||  
Db 330 GLFGAAGFIENGMEGMDGMYG 352

RESULT 2  
HEMA\_IADH2 STANDARD: PRT: 550 AA.  
ID HEMA\_IADH2 084011:  
AC P12583; 084011:  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hokkaido/8/80).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A virus.  
OC Influenza A virus.  
OX NCBI\_TaxID=11358;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87265458; PubMed=2440178;  
RA Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;  
RT "Antigenic and genetic conservation of H3 influenza virus in wild  
ducks."  
RL Virology 159:109-119(1987).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M16738; AAA43144.1: -  
DR PIR: B27813; HMTV80.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR Prodom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT K -> N (IN PIR DATA BANK).  
SQ SEQUENCE 550 AA; 61659 MW; A107023AC9CC353 CRC64;

Query Match 55.18; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAAGFIENGMEGMDGMYG 24  
|||||  
Db 330 GLFGAAGFIENGMEGMDGMYG 352

RESULT 3  
HEMA\_IADH3

ID HEMA\_IADH3 STANDARD: PRT: 550 AA.  
AC P12584; 084012; 089793;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hokkaido/33/80).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A virus.  
OC Influenza A virus.  
OX NCBI\_TaxID=11359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87265458; PubMed=2440178;  
RA Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;  
RT "Antigenic and genetic conservation of H3 influenza virus in wild  
ducks."  
RL Virology 159:109-119(1987).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M16739; AAA43145.1: -  
DR PIR: C27813; HMTV33.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR Prodom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA; 61577 MW; 6C30BF67CFCB7DE CRC64;

Query Match 55.18; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAAGFIENGMEGMDGMYG 24  
|||||  
Db 330 GLFGAAGFIENGMEGMDGMYG 352

RESULT 4  
HEMA\_IADH4 STANDARD: PRT: 550 AA.  
ID HEMA\_IADH4 084013; 084014;  
AC P12585; 084013; 084014;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hokkaido/7/82).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

```
CC Influenza virus A and B group; Influenza A viruses;
CC Influenza A virus.
OX NCBI_TaxID=11360;
RN (1)
RP MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
   ducks."
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC EMBL: M16740; AAA43146.1;
CC PIR: D27813; HMIY89.
CC InterPro: IPR001364; Hemagglutn.
CC Pfam: PF00509; Hemagglutinin; 1.
CC ProDom: PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; glycoprotein.
CC NON_TER 1 1
CC CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
CC CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 550 AA; 61664 MW; A16B2CF8CBBD9D0 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFENGMEGMDGMYG 24
DB 330 GLFGAIGFENGMEGMDGMYG 352

RESULT 5
ID HEMA_IADH5 STANDARD; PRT; 550 AA.
AC P12586; Q84015; Q84016;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
   Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/21/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX NCBI_TaxID=11361;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
   ducks."
RL Virology 159:109-119(1987).
```

```
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC EMBL: M16741; AAA43147.1;
CC PIR: E27813; HMIY21.
CC HSSP: P03437; 5HMG.
CC InterPro: IPR001364; Hemagglutn.
CC Pfam: PF00509; Hemagglutinin; 1.
CC ProDom: PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; glycoprotein.
CC NON_TER 1 1
CC CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
CC CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 7 7 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CONFLICT 178 179 YV -> VI (IN PIR DATA BANK).
CC SEQUENCE 550 AA; 61856 MW; 48401C867A15BF8C CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFENGMEGMDGMYG 24
DB 330 GLFGAIGFENGMEGMDGMYG 352

RESULT 6
ID HEMA_IADH6 STANDARD; PRT; 550 AA.
AC P12587; Q84017;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
   Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX NCBI_TaxID=11362;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
   ducks."
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M16742; AAA43148.1; -  
DR PIR: F27813; HMIY98.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KM Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 328  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 8 8 Y -> N (IN PIR DATA BANK).  
SO SEQUENCE 550 AA; 61711 MW; 67BCD85F44736CFE CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFLENGEGMIDGMYG 24  
DB 330 GLFGAIGFLENGEGMIDGMYG 352

## RESULT 7

HEMA\_IADH7 STANDARD: PRT; 550 AA.  
ID HEMA\_IADH7 Q84018; Q89470;

DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain] (Fragment).

OS Influenza A virus (strain A/Duck/Hokkaido/10/85).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OX NCBI\_TaxID-11363;

RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE-87265458; PubMed-2440178;  
RA Kida H., Kawacka Y., Naeve C.W., Webster R.G.;  
RT "Antigenic and genetic conservation of H3 influenza virus in wild  
RT ducks."

RL Virology 159:109-119(1987).

CC - FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.

CC - SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC - SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M16743; AAA43149.1; -  
DR PIR: G27813; HMIY15.

DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KM Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 328  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SO SEQUENCE 550 AA; 61761 MW; 6EF81793281D53EB CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFLENGEGMIDGMYG 24  
DB 330 GLFGAIGFLENGEGMIDGMYG 352

RESULT 8  
HEMA\_IADHK STANDARD: PRT; 550 AA.  
ID HEMA\_IADHK P43257;

DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain] (Fragment).

OS Influenza A virus (strain A/Duck/Hong Kong/7/75).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OX NCBI\_TaxID-11364;

RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE-9134191; PubMed-1875195;  
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
RT "Molecular evidence for a role of domestic ducks in the introduction  
RT of avian H3 influenza viruses to pigs in southern China, where the  
RT A/Hong Kong/68 (H3N2) strain emerged."

RL J. Gen. Virol. 72:2007-2010(1991).

CC - FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.

CC - SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC - SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: D00929; BAA00769.1; -  
DR HSSP: P03437; 5HM0.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR ProDom: PD000225; Hemagglutn; 1.

KM Envelope protein; Hemagglutinin; Glycoprotein.

FT NON\_TER 1 328  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SO SEQUENCE 550 AA; 61761 MW; 6EF81793281D53EB CRC64;

FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SO SEQUENCE 550 AA; 61761 MW; 6EF81793281D53EB CRC64;

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA: 61549 MW: 864639B829FE1BA9 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAFGIENGMEGMDWG 24  
Db 330 GLFGAIAFGIENGMEGMDWG 352

RESULT 9  
HEMA\_IADHL STANDARD: PRT: 550 AA.  
ID HEMA\_IADHL  
AC P43258;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hong Kong/64/76).  
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=45412;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91341491; Pubmed=1875195;  
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
RT "Molecular evidence for a role of domestic ducks in the introduction  
of avian H3 influenza viruses to pigs in southern China, where the  
A/Hong Kong/68 (H3N2) strain emerged.";  
RT J. Gen. Virol. 72:2007-2010(1991)  
RL J. Gen. Virol. 72:2007-2010(1991)  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D00931; BAA00771.1; -  
DR HSSP: P03437; 2HMG.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin: 1.  
DR ProDom: PD000225; Hemagglutn: 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON-TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA: 61718 MW: A351C56789E4BE9A CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GLFGAIAFGIENGMEGMDWG 24  
Db 330 GLFGAIAFGIENGMEGMDWG 352

RESULT 10  
HEMA\_IAGHK STANDARD: PRT: 550 AA.  
ID HEMA\_IAGHK  
AC P43260;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Goose/Hong Kong/10/76).  
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=45414;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91341491; Pubmed=1875195;  
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
RT "Molecular evidence for a role of domestic ducks in the introduction  
of avian H3 influenza viruses to pigs in southern China, where the  
A/Hong Kong/68 (H3N2) strain emerged.";  
RT J. Gen. Virol. 72:2007-2010(1991)  
RL J. Gen. Virol. 72:2007-2010(1991)  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D00930; BAA00770.1; -  
DR HSSP: P03437; 5HMG.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin: 1.  
DR ProDom: PD000225; Hemagglutn: 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON-TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA: 61676 MW: 9A1E094DA28BACD2 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAFGIENGMEGMDWG 24  
Db 330 GLFGAIAFGIENGMEGMDWG 352

RESULT 11  
HEMA\_IAGH2

```
ID HEMA_IATZ2 STANDARD: PRT: 550 AA.
AC P1133; 084019; 084020;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1989 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/81/78).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX Influenza A virus.
RN NCB1_TaxID=11497;
RX MEDLINE=8101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.";
RL Virology 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M19057; AAA43212.1; -.
CC PIR: B29971; HMIVS3.
CC InterPro: IPR001364; Hemagglutn.
CC Pfam: PF00509; Hemagglutinin; 1.
CC ProDom: PD000225; Hemagglutn; 1.
CC Hemagglutinin; Envelope protein; Glycoprotein.
CC NON_TER 1 1
CC CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
CC FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
CC FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 550 AA; 61437 MW; 1F2A7E758C531CE8 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEGAIGFIENGMEGMDGMYG 24
DB 330 GLEGAIGFIENGMEGMDGMYG 352

RESULT 12
HEMA_IATZ3 STANDARD: PRT: 550 AA.
AC P1134; 084025; 084026;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1989 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/126/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
```

```
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCB1_TaxID=11498;
RN NCB1_TaxID=11498;
RX MEDLINE=8101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.";
RL Virology 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M19056; AAA43211.1; ALT_TERM.
CC PIR: A29971; HMIVS2.
CC HSSP: P03437; 2HMG.
CC InterPro: IPR001364; Hemagglutn.
CC Pfam: PF00509; Hemagglutinin; 1.
CC ProDom: PD000225; Hemagglutn; 1.
CC Hemagglutinin; Envelope protein; Glycoprotein.
CC NON_TER 1 1
CC CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
CC FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
CC FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 550 AA; 61580 MW; 991F6DB8C02F24F2 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEGAIGFIENGMEGMDGMYG 24
DB 330 GLEGAIGFIENGMEGMDGMYG 352

HEMA_IATZ3 STANDARD: PRT: 566 AA.
AC P03437;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Alchi/2/68).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX Influenza A virus.
RN NCB1_TaxID=150147;
RX MEDLINE=80254693; PubMed=7402351;
RA Verhoeven M., Pang R., Min Jou W., Devos R., Huylebroeck D., Saman E., Fiers W.;
RT "Antigenic drift between the haemagglutinin of the Hong Kong
```





```
FT STRAND 310 311
FT STRAND 318 320
FT STRAND 320 324
FT STRAND 321 324
FT STRAND 331 333
FT STRAND 337 337
FT STRAND 347 348
FT STRAND 350 350
FT STRAND 351 351
FT STRAND 352 354
FT STRAND 355 355
FT STRAND 359 359
FT STRAND 360 361
FT STRAND 367 372
FT STRAND 374 375
FT STRAND 378 382
FT STRAND 383 401
FT STRAND 406 407
FT STRAND 421 474
FT STRAND 475 477
FT STRAND 482 485
FT STRAND 491 498
FT STRAND 499 500
FT STRAND 505 515
SQ SEQUENCE 566 AA: 63415 MW: E395659C23CAFECA CRC64;
```

Query Match 55.1%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 2.4e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 2 GLFGAAGFIENGMEGMDWYG 24
Db 346 GLFGAAGFIENGMEGMDWYG 368
|||||
```

```
RESULT 14
HEMA_IAD3 STANDARD: PRT; 566 AA.
ID HEMA_IAD3
AC P26134;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Duck/Alberta/78/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11348;
RN 11)
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; Pubmed=1731092;
RA Bean W.J., Schell M., Katz J., Kawoka Y., Naeve C., Gorman O.,
RA Webster R.G.;
RT Evolution of the H3 influenza virus hemagglutinin from human and
RT nonhuman hosts.
RL J. Virol. 66:1129-1138(1992).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M73771; -; NOT_ANNOTATED_CDS.
DR
```

```
DR HSSP; P03437; SHMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn.1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344
FT CHAIN 346 566
FT CARBOHYD 23 23
FT CARBOHYD 24 24
FT CARBOHYD 38 38
FT CARBOHYD 54 54
FT CARBOHYD 181 181
FT CARBOHYD 301 301
FT CARBOHYD 499 499
SQ SEQUENCE 566 AA: 63534 MW: FE19AB6FF9415B89 CRC64;
```

Query Match 55.1%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 2.4e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 2 GLFGAAGFIENGMEGMDWYG 24
Db 346 GLFGAAGFIENGMEGMDWYG 368
|||||
```

```
RESULT 15
HEMA_IAD3 STANDARD: PRT; 566 AA.
ID HEMA_IAD3
AC P03442;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Duck/Ukraine/1/63).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11374;
RN 11)
RP SEQUENCE FROM N.A.
RX MEDLINE=82025542; Pubmed=6169439;
RA Tang R., Min Jou W., Huylebroeck D., Devos R., Fiers W.;
RA "Complete structure of A/Duck/Ukraine/63 influenza hemagglutinin
RT gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza
RT hemagglutinin".
RL Cell 25:315-323(1981).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: V01087; CAA24271.1; -
CC PIR: A04053; HMTVD.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn.1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
```

FT	CHAIN	17	344	HEMAGGLUTININ HA1 CHAIN.
FT	CHAIN	346	566	HEMAGGLUTININ HA2 CHAIN.
FT	CARBOHYD	24	24	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	38	38	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	54	54	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	97	97	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	301	301	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	499	499	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	566 AA;	63530 MW;	E70F87F0AE1178F4 CRC64;

Query Match 55.1%: Score 134; DB 1; Length 566;  
 Best Local Similarity 100.0%: Pred. No. 2.4e-07;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 2 GLFGAIGFIENGWEGMIDGMYG 24  
 |||||  
 Db 346 GLFGAIGFIENGWEGMIDGMYG 368

Search completed: July 1, 2002, 06:31:35  
 Job time: 672 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 1, 2002, 06:30:59 ; Search time 83.98 Seconds

(without alignments)  
90.638 Million cell updates/sec

Title: US-09-461-684-5  
Perfect score: 243  
Sequence: 1 CGLFGAIGFIENGWEGMID.....KKKKKKKKKKKKKKKKKK 44

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	55.1	566	12	Q9DHG0
2	134	55.1	566	12	Q67125
3	134	55.1	566	12	Q67126
4	134	55.1	566	12	Q67132
5	134	55.1	566	12	Q98052
6	134	55.1	566	12	Q91MA7
7	134	55.1	566	12	Q91OM5
8	133	54.7	301	12	Q9DXE3
9	132	54.3	550	12	Q82499
10	132	54.3	550	12	Q82499
11	132	54.3	550	12	Q82499
12	132	54.3	566	12	Q82496
13	132	54.3	571	12	Q03909
14	131	53.9	109	12	Q67050
15	131	53.9	109	12	Q67051
16	131	53.9	109	12	Q67052

17	131	53.9	109	12	Q67053	067053	Influenza a
18	131	53.9	362	12	Q82513	082513	Influenza a
19	131	53.9	362	12	Q82517	082517	Influenza a
20	131	53.9	362	12	Q84174	084174	Influenza a
21	131	53.9	362	12	Q9QKD3	09QKD3	Influenza a
22	131	53.9	362	12	Q9QKD2	09QKD2	Influenza a
23	131	53.9	362	12	Q9QKD1	09QKD1	Influenza a
24	131	53.9	365	12	Q9DL25	09DL25	Influenza a
25	131	53.9	367	12	Q9DL22	09DL22	Influenza a
26	131	53.9	368	12	Q9DL29	09DL29	Influenza a
27	131	53.9	369	12	Q9DL26	09DL26	Influenza a
28	131	53.9	369	12	Q9DL06	09DL06	Influenza a
29	131	53.9	369	12	P87689	P87689	Influenza a
30	131	53.9	371	12	Q9DL24	09DL24	Influenza a
31	131	53.9	371	12	P87685	P87685	Influenza a
32	131	53.9	373	12	Q9DL20	09DL20	Influenza a
33	131	53.9	374	12	Q9DL21	09DL21	Influenza a
34	131	53.9	375	12	Q9DL27	09DL27	Influenza a
35	131	53.9	375	12	Q9DL05	09DL05	Influenza a
36	131	53.9	376	12	Q9DL30	09DL30	Influenza a
37	131	53.9	376	12	Q9DL04	09DL04	Influenza a
38	131	53.9	377	12	Q9E7P6	Q9E7P6	Influenza a
39	131	53.9	382	12	Q9DL03	09DL03	Influenza a
40	131	53.9	408	12	Q9E7P5	Q9E7P5	Influenza a
41	131	53.9	409	12	Q9Q0L5	09Q0L5	Influenza a
42	131	53.9	416	12	Q9Q0C4	09Q0C4	Influenza a
43	131	53.9	429	12	Q9Q0L4	09Q0L4	Influenza a
44	131	53.9	438	12	Q9Q0L3	09Q0L3	Influenza a
45	131	53.9	467	12	Q9E312	09E312	Influenza a

## ALIGNMENTS

RESULT 1  
ID Q9DHG0 PRELIMINARY: PRT: 566 AA.  
AC Q9DHG0;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HAEMAGGLUTININ PRECURSOR.  
OS Influenza A virus H3N2.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae.  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_Taxid=41857;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLONE 7A (H3N2);  
RA Moshin M.A., Morris S.J., Smith H., Sweet C.;  
RT "Influenza virus-induced apoptosis: a dual role for viral neuraminidase".  
RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RL -1- FUNCTION: HAEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HAEMAGGLUTININ FAMILY.  
CC EMBL: AJ289703; CAC18525.1; -  
DR HSSP: P03437; 2YIU  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HAEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutin.1.  
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.  
FT SIGNAL 16 POTENTIAL  
SQ SSEQUENCE 566 AA; 63356 MW; OBA681929300F72F CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 46-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY      2  GLFGAIAGFLENGMEGMDGWC 24
|||||
Db      346  GLFGAIAGFLENGMEGMDGWC 368

RESULT  2
O67125      PRELIMINARY:      PRT:      566 AA.
AC 067125:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMAGGLUTININ.
GN HA.
OS Influenza A virus.
OC Viruses: ssRNA negative-strand viruses: Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/SEAL/MA/3911/92;
RX MEDLINE=95146951; PubMed=7844533;
RA Callan R.J., Early G., Kida H., Hinshaw V.S.;
RT "The appearance of H3 influenza viruses in seals.";
RL J. Gen. Virol. 76:199-203(1995).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: L31949; AAA64229.1;
DR HSSP: P03437; 2VIU.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn.1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
KW SEQUENCE 566 AA; 63456 MW; AE556302A9EB99F CRC64;
SQ

Query Match      55.1%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2  GLFGAIAGFLENGMEGMDGWC 24
|||||
Db      346  GLFGAIAGFLENGMEGMDGWC 368

RESULT  3
O67126      PRELIMINARY:      PRT:      566 AA.
AC 067126:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMAGGLUTININ.
GN HA.
OS Influenza A virus.
OC Viruses: ssRNA negative-strand viruses: Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/SEAL/MA/3984/92;
RX MEDLINE=95146951; PubMed=7844533;
RA Callan R.J., Early G., Kida H., Hinshaw V.S.;
RT "The appearance of H3 influenza viruses in seals.";
RL J. Gen. Virol. 76:199-203(1995).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
DR EMBL: M55059; AAA43239.1;
DR HSSP: P03437; 1HCE.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn.1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
KW CHAIN 1 344 HEMAGGLUTININ.
FT CHAIN 346 566 NEURAMINIDASE.
SQ SEQUENCE 566 AA; 63441 MW; ESD1B97DF6FECA CRC64;

Query Match      55.1%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2  GLFGAIAGFLENGMEGMDGWC 24
|||||
Db      346  GLFGAIAGFLENGMEGMDGWC 368

RESULT  5
O98052      PRELIMINARY:      PRT:      566 AA.
AC 098052:
DT 01-FEB-1997 (TREMBLrel. 02, Created)
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: L32024; AAA64228.1;
DR HSSP: P03437; 2VIU.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn.1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63441 MW; 590576CB4CE7D08 CRC64;

Query Match      55.1%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2  GLFGAIAGFLENGMEGMDGWC 24
|||||
Db      346  GLFGAIAGFLENGMEGMDGWC 368

RESULT  4
O67132      PRELIMINARY:      PRT:      566 AA.
AC 067132:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMAGGLUTININ.
GN HA.
OS Influenza A virus (strain A/Aichi/2/68).
OC Viruses: ssRNA negative-strand viruses: Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=150147;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/AICHI/2/68;
RX Min J.W., Verhoeven M., Fang R.-X., Devos R., Huylebroeck D.,
RA Fiers W.;
RT "Shift and drift in influenza viruses.";
RL (in) Carille M.J., Collins J.F., Moseley B.E. B. (eds.);
RL Symposium of the Society for General Microbiology, pp.285-311,
RL Cambridge University Press, New York (1981).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: M55059; AAA43239.1;
DR HSSP: P03437; 1HCE.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn.1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
KW CHAIN 1 344 HEMAGGLUTININ.
FT CHAIN 346 566 NEURAMINIDASE.
SQ SEQUENCE 566 AA; 63441 MW; ESD1B97DF6FECA CRC64;

Query Match      55.1%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2  GLFGAIAGFLENGMEGMDGWC 24
|||||
Db      346  GLFGAIAGFLENGMEGMDGWC 368

RESULT  5
O98052      PRELIMINARY:      PRT:      566 AA.
AC 098052:
DT 01-FEB-1997 (TREMBLrel. 02, Created)
```

DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEMAGGLUTININ PRECURSOR (FRAGMENT).  
OS Influenza A virus.  
CC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID=11320;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=81053698; PubMed=6253883;  
RA Both G.W., Sleight M.J.;  
RT "Complete nucleotide sequence of the haemagglutinin gene from a human  
influenza virus of the Hong Kong subtype.";  
RL Nucleic Acids Res. 8:2561-2575(1980).  
RN (12)  
RP SEQUENCE OF 17-344 FROM N.A.  
RX MEDLINE=81194918; PubMed=6164798;  
RA Sleight M.J., Both G.W., Underwood P.A., Bender V.J.;  
RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza  
subtype: Correlation of amino acid changes with alterations in viral  
antigenicity.";  
RL J. Virol. 37:845-853(1981).  
RN (13)  
RP SEQUENCE OF 17-566 FROM N.A.  
RX MEDLINE=82033276; PubMed=6169843;  
RA Both G.W., Sleight M.J.;  
RT "Conservation and variation in the hemagglutinins of Hong Kong subtype  
influenza viruses during antigenic drift.";  
RL J. Virol. 39:845-853(1981).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: J02135; AAA43189.1; -  
DR HSSP: P03437; IHGE.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; Hemagglutn.1.  
DR PRODOM: PD000225; Hemagglutn.1.  
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.  
FT SIGNAL 1 16 POTENTIAL.  
FT CHAIN 17 344 POTENTIAL.  
FT CHAIN 346 566 POTENTIAL.  
SQ SEQUENCE 566 AA: 63414 MW: C447FD465BEAFCE9 CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMIDGWG 24  
DB 346 GLFGAIAGFIENGWEGMIDGWG 368

RESULT 6  
O91MA7 PRELIMINARY; PRT: 566 AA.  
AC O91MA7;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE HEMAGGLUTININ.  
OS Influenza A virus (A/Hong Kong/1/68(H3N2)).  
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=108859;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX STRAIN=A/HONG KONG/1/68(H3N2);  
RA MEDLINE=21287244; PubMed=11371620;  
RL Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;

RT "Pattern of mutation in the genome of influenza A virus on adaptation  
to increased virulence in the mouse lung: Identification of functional  
themes.";  
RT Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).  
DR EMBL: AF348176; AAK51718.1; -  
SQ SEQUENCE 566 AA: 63387 MW: 01BBD0465BE158E1 CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMIDGWG 24  
DB 346 GLFGAIAGFIENGWEGMIDGWG 368

RESULT 7  
O91OM5 PRELIMINARY; PRT: 566 AA.  
AC O91OM5;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE HEMAGGLUTININ.  
OS Influenza A virus (A/Hong Kong/1/68(H3N2)).  
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=108859;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX STRAIN=A/HONG KONG/1/68(H3N2);  
RA MEDLINE=21287244; PubMed=11371620;  
RL Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;  
RT "Pattern of mutation in the genome of influenza A virus on adaptation  
to increased virulence in the mouse lung: Identification of functional  
themes.";  
RT Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).  
DR EMBL: AF348179; AAK51721.1; -  
DR EMBL: AF348177; AAK51719.1; -  
DR EMBL: AF348178; AAK51720.1; -  
SQ SEQUENCE 566 AA: 63530 MW: 7CB9F5B8F1E6E9F4 CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMIDGWG 24  
DB 346 GLFGAIAGFIENGWEGMIDGWG 368

RESULT 8  
O9DXE3 PRELIMINARY; PRT: 301 AA.  
AC O9DXE3;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE HEMAGGLUTININ (FRAGMENT).  
CN HA1.  
OS Influenza A virus (A/Shorebird/Taiwan/31-4/99).  
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=140665;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX STRAIN=A/SHOREBIRD/TAIWAN/31-4/99;  
RA Lee M.S., Cheng P.C., Shien J.H., Cheng M.C., Lee L.H., Shieh H.K.;  
RT "Identification and subtyping of avian influenza virus by reverse  
transcription-polymerase chain reaction.";  
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.

```
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: AF311730; AAC33016.1; -.
DR InterPro: IPR001364; Hemagglutn.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 301
FT SEQUENCE 301 AA; 32701 MW; 62A403758B764D57 CRC64;

OY Query Match 54.3%; Score 132; DB 12; Length 301;
Best Local Similarity 95.7%; Pred. No. 2.7e-08;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWY 24
DB 250 GLFGAIGFIENGWEGMIDGWY 272

RESULT 9
ID 082499 PRELIMINARY; PRT; 550 AA.
AC 082499;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMAGGLUTININS HA1 AND HA2 (FRAGMENT).
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxId=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82/BS (H3N2);
RA Hartley C.A., Ward A.C., Anders E.M.;
RT "Virulence of Influenza virus for mice is associated with loss of
RT oligosaccharide from the hemagglutinin molecule.";
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC EMBL: U08859; AAA18782.1; -.
DR HSP: P03437; 2YIU.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 328
FT CHAIN 1 330 HA1.
FT CHAIN 330 550 HA2.
FT SEQUENCE 550 AA; 61772 MW; 50BD62B6BE11FD8 CRC64;

OY Query Match 54.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 6.7e-08;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWY 24
DB 330 GLFGAIGFIENGWEGMIDGWY 352

RESULT 10
ID 082753 PRELIMINARY; PRT; 550 AA.
```

```
AC 082753;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HAEMAGGLUTININ (FRAGMENT).
OS Influenza virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC unclassified Orthomyxoviridae.
OX NCBI_TaxId=11309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82/BS/ML10 (H3N2);
RX MEDLINE=97300854; PubMed=9155874;
RA Hartley C.A., Reading P.C., Ward A.C., Anders E.M.;
RT "Changes in the hemagglutinin molecule of Influenza type A (H3N2)
RT virus associated with increased virulence for mice.";
RL Arch. Virol. 142:75-86(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82/BS/ML10 (H3N2);
RX MEDLINE=97456249; PubMed=9311563;
RA Ward A.C.;
RT "Virulence of Influenza A virus for mouse lung.";
RL Virus genes 14:187-194(1997).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC EMBL: U08905; AAC79579.1; -.
DR HSP: P03437; 2YIU.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 328
FT CHAIN 1 330 HAEMAGGLUTININ HA1.
FT CHAIN 330 550 HAEMAGGLUTININ HA2.
FT SEQUENCE 550 AA; 61745 MW; 692A49DE678AC4BC CRC64;

OY Query Match 54.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 6.7e-08;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWY 24
DB 330 GLFGAIGFIENGWEGMIDGWY 352

RESULT 11
ID 082498 PRELIMINARY; PRT; 550 AA.
AC 082498;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMAGGLUTININS HA1 AND HA2 (FRAGMENT).
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxId=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82 (H3N2);
RA Hartley C.A., Ward A.C., Anders E.M.;
RT "Virulence of Influenza virus for mice is associated with loss of
RT oligosaccharide from the hemagglutinin molecule.";
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82 (H3N2);
```

RX MEDLINE-88185444; PubMed-3356226;  
RA Nakajima S., Takeuchi Y., Nakajima K.;  
RT "Location on the evolutionary tree of influenza H3 haemagglutinin  
genes of Japanese strains isolated during the 1985-6 season.";  
RL Epidemiol. Infect. 100:301-310(1988).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: U08858; AAA18781.1; -.  
DR HSSP: P03437; 2V1U.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutn.1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HA1.  
FT CHAIN 330 550 HA2.  
SQ SEQUENCE 550 AA; 61802 MW; 1144131CE5A1P6A CRC64;

Query Match 54.3%; Score 132; DB 12; Length 550;  
Best Local Similarity 95.7%; Pred. No. 6.7e-08;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGMEGMDGMYG 24  
DB 330 GIFGAIAGFIENGMEGMDGMYG 352  
1:|||||

RESULT 12  
ID 082496 PRELIMINARY; PRT; 566 AA.  
AC 082496;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEMAGGLUTININ.  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID-11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/SWINE/ANGE-GARDIEN/150/90(H3N2);  
RX MEDLINE-95205091; PubMed-7897358;  
RA Bikour M.H., Frost E.H., Deslandes S., Talbot B., Weber J.M.,  
RA Elazhary Y.;  
RT "Recent H3N2 swine influenza virus with haemagglutinin and  
nucleoprotein genes similar to 1975 human strains.";  
RL J. Gen. Virol. 76:697-703(1995).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY)  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: U07146; AAA85781.1; -.  
DR HSSP: P03437; 2V1U.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutn.1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
SQ SEQUENCE 566 AA; 63362 MW; 1FB8319A567E2FEF CRC64;

Query Match 54.3%; Score 132; DB 12; Length 566;  
Best Local Similarity 95.7%; Pred. No. 6.9e-08;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGMEGMDGMYG 24  
DB 346 GIFGAIAGFIENGMEGMDGMYG 368  
1:|||||

RESULT 13  
ID 003909 PRELIMINARY; PRT; 571 AA.  
AC 003909;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEMAGGLUTININ PRECURSOR.  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID-11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Guo Y., Wang M., Kawaka Y., Gorman O.T., Ito T., Webster R.G.;  
RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION.  
DR EMBL: M65018; AAA43151.1; -.  
DR HSSP: P03437; 2V1U.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutn.1.  
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 349 HA1 CHAIN.  
FT CHAIN 350 571 HA2 CHAIN.  
FT DISULFID 36 487 BY SIMILARITY.  
FT DISULFID 73 298 BY SIMILARITY.  
FT DISULFID 85 297 BY SIMILARITY.  
FT DISULFID 160 494 BY SIMILARITY.  
SQ SEQUENCE 571 AA; 64104 MW; 718DA0F291CE349 CRC64;

Query Match 54.3%; Score 132; DB 12; Length 571;  
Best Local Similarity 95.7%; Pred. No. 7e-08;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGMEGMDGMYG 24  
DB 351 GIFGAIAGFIENGMEGMDGMYG 373  
1:|||||

RESULT 14  
ID 067050 PRELIMINARY; PRT; 109 AA.  
AC 067050;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEMAGGLUTININ (FRAGMENT).  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID-11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/FUKUOKA/C29/85 (H3N2);  
RX MEDLINE-81030852; PubMed-7421990;  
RA Gething M.-J., Bye J., Skehel J., Waterfield M.;  
RT "Cloning and dna sequence of double-stranded copies of haemagglutinin  
genes from h2 and h3 strains elucidates antigenic shift and drift in  
human influenza virus.";  
RL Nature 287:301-306(1980).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN-A/FUKUOKA/C29/85 (H3N2):  
RX MEDLINE-93233219; PubMed-7682624;  
RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;  
RT "A common neutralizing epitope conserved between the hemagglutinins of  
influenza A virus H1 and H2 strains.";  
RL J. Virol. 67:2552-2558(1993).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: D13581; BAA02776.1; -.  
DR HSSP: P03437; 1HTM.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutn; 1.  
KM Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON\_TER 1 109 1  
FT SEQUENCE 109 AA; 12305 MW; 17EC66753C48672F CRC64;

Query Match 53.9%; Score 131; DB 12; Length 109;  
Best Local Similarity 91.3%; Pred. No. 1.5e-08;  
Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIEGMEGMDGWTG 24  
1:|||||  
DB 40 GIFGAIAGFIEGMEGMDGWTG 62

RESULT 15  
067051  
ID 067051 PRELIMINARY; PRT; 109 AA.  
AC 067051.  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEMAGGLUTININ (FRAGMENT).  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID-11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/SICHUAN/2/87 (H3N2):  
RX MEDLINE-81030852; PubMed-7421990;  
RA Gething M.-J., Bye J., Skehel J., Waterfield M.;  
RT "Cloning and dna sequence of double-stranded copies of haemagglutinin  
genes from h2 and h3 strains elucidates antigenic shift and drift in  
human influenza virus.";  
RL Nature 287:301-306(1980).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/SICHUAN/2/87 (H3N2):  
RX MEDLINE-93233219; PubMed-7682624;  
RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;  
RT "A common neutralizing epitope conserved between the hemagglutinins of  
influenza A virus H1 and H2 strains.";  
RL J. Virol. 67:2552-2558(1993).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: D13582; BAA02777.1; -.  
DR HSRP: P03437; 1HGE.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR PRINTS: PR00329; HEMAGGLUTN12.

DR ProDom: PD000225; Hemagglutn; 1.  
KM Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON\_TER 1 109 1  
FT SEQUENCE 109 AA; 12293 MW; 17EC66752DB8672F CRC64;

Query Match 53.9%; Score 131; DB 12; Length 109;  
Best Local Similarity 91.3%; Pred. No. 1.5e-08;  
Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIEGMEGMDGWTG 24  
1:|||||  
DB 40 GIFGAIAGFIEGMEGMDGWTG 62

Search completed: July 1, 2002, 06:30:59  
Job time: 691 sec



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:25 ; Search time 98.97 Seconds  
(Without alignments)  
49.381 Million cell updates/sec

Title: US-09-461-684-5

Perfect score: 243  
Sequence: 1 CGLFGAINGFIENGMEGMID.....KKKKKKKKKKKKKKKKKK 44

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

## Database :

Listing first 45 summaries

```

1: A_Geneseq_032802.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	243	100.0	44	21	AA13784
2	192	79.0	40	21	AA151431
3	143	58.8	24	21	AA13782
4	143	58.8	26	13	AA12604
5	143	58.8	26	18	AA134269
6	143	58.8	26	22	AA170090
7	134	55.1	23	17	AA129226
8	134	55.1	25	18	AA198016
9	134	55.1	26	13	AA126005
10	134	55.1	26	18	AA134270
11	134	55.1	26	22	AA170091

12	134	55.1	28	17	AA198024
13	134	55.1	30	21	AA198680
14	134	55.1	30	22	AA1801088
15	134	55.1	30	22	AA1802537
16	134	55.1	30	22	AA1803634
17	134	55.1	30	22	AA1808081
18	134	55.1	175	15	AA185797
19	134	55.1	223	22	AA14032
20	134	55.1	566	15	AA183590
21	134	55.1	685	22	AA187893
22	134	55.1	29	20	AA129710
23	132	54.3	221	14	AA188864
24	132	54.3	221	14	AA188865
25	132	54.3	221	15	AA180220
26	132	54.3	221	15	AA180221
27	132	54.3	221	15	AA180221
28	132	54.3	306	14	AA188867
29	132	54.3	306	15	AA180194
30	132	54.3	386	5	AA140615
31	131	53.9	565	8	AA170711
32	131	53.9	565	11	AA180493
33	131	53.9	565	19	AA144946
34	131	53.9	565	21	AA170056
35	131	53.9	565	21	AA170057
36	131	53.9	566	19	AA188406
37	131	53.9	570	8	AA170710
38	131	53.9	570	11	AA180494
39	131	53.9	570	18	AA180169
40	131	53.9	570	20	AA180491
41	131	53.9	571	22	AA180491
42	131	53.9	571	18	AA180167
43	131	53.9	571	18	AA180167
44	131	53.9	571	20	AA175445
45	131	53.9	571	20	AA175448

## ALIGNMENTS

RESULT 1	AA13784	standard; peptide: 44 AA.
XX	AA13784:	
AC	AA13784:	
XX	10-NOV-2000	(first entry)
XX		
DE	Soluble tandem HA/ PK peptide conjugate.	
KW	pk peptide; cytosolic; vaccine; cytotoxic T cell; CTL; immunotherapy;	
KW	major histocompatibility complex class I; MHC class I; antigen; tumour;	
KW	prostate; breast; multiple myeloma; HA peptide.	
XX	Unidentified.	
OS		
XX	WO200035949-A1.	
PN		
XX	22-JUN-2000.	
PD		
XX	14-DEC-1999;	99MO-US29724.
PF		
XX	14-DEC-1998;	98US-0112324.
PR		
XX	(DEND-) DENDREON CORP.	
PA		
XX	Iaus R, Hakim I, Vidovic D;	
PI		
XX	WPI, 2000-442365/38.	
DR		
XX		
PT	Antigens modified by the covalent addition of a peptide that	
PT	facilitates entry into antigen presenting cells, useful for producing	
PT	compositions for immunizing against tumors and pathogens -	
XX		

PS Claim 2; Page 26; 34pp: English.

XX CC The present invention relates to compositions of modified soluble protein  
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)  
CC response i.e. a major histocompatibility complex (MHC) class I molecule  
CC response. The protein antigen is modified by the covalent addition of a  
CC peptide sequence which facilitate entry of the antigen into antigen  
CC presenting cells (APCs). The present sequence is one such peptide  
CC sequence which can be used to modify the soluble antigens. The present  
CC sequence is tandem HA/ PK peptide conjugate. The modified antigen  
CC composition may be used for immunising against, or treating a tumour e.g.  
CC prostate and breast carcinoma or multiple myeloma, or pathogen in  
CC mammals.

XX SQ Sequence 44 AA:

Query Match 100.0%; Score 243; DB 21; Length 44;  
Best Local Similarity 100.0%; Pred. No. 7.4e-23;  
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAIAFGIENGWEGMIDGMYGKKKKKKKKKKKKKKKKKKKK 44  
Db 1 cglfgaiafgiengwegmidgwygkkkkkkkkkkkkkkkkkkkk 44

RESULT 2  
AAM51431  
ID AAM51431 standard; peptide: 40 AA.

XX AC AAM51431:  
XX DT 28-JAN-2002 (first entry)  
XX DE Peptide K14 Fuso Influenza haemagglutinin fusogenic peptide.  
XX KM Gene therapy vector; cell entry; intracellular trafficking;  
XX KW gene expression; Influenza: haemagglutinin.  
XX OS Synthetic.  
XX PN WO200149324-A2.  
XX PD 12-JUL-2001.  
XX PF 28-DEC-2000; 2000WO-EPI3300.  
XX PR 30-DEC-1999; 99US-0475305.  
XX PA (NOVS ) NOVARTIS AG.  
XX PA (NOVS ) NOVARTIS-ERRINDUNGEN VERN GES MBH.  
XX PI Woodle M, Cheng C, Puthupparampil S, Subramanian K, Tilmas R;  
XX PI Yang J, Frel J, Mett H, Stanek J;  
XX DR WPI: 2001-602251/68.  
XX PT Non-naturally occurring gene therapy vector useful for gene therapy,  
XX PT comprises an inner shell having a core complex containing a nucleic  
XX PT acid and at least one complex forming reagent -  
XX PS Example 46; Page 99; 178pp: English.

XX CC The invention relates to a non-naturally occurring gene therapy vector,  
XX CC comprising an inner shell having a core complex containing a nucleic acid  
XX CC and at least one complex forming reagent. The vectors are stable having  
XX CC an improved outer steric layer that provides enhanced target specificity,  
XX CC in vivo and colloidal stability. The vectors are relatively homogeneous  
XX CC and comprise chemically defined species. The vectors demonstrate improved  
XX CC Cq11 entry and intracellular trafficking, permitting enhanced nucleic  
XX CC acid therapeutic activity such as gene expression. The present sequence  
XX CC is that of the peptide K14 Fuso, comprising a fusogenic peptide derived  
XX CC from influenza haemagglutinin and peptide K14, useful in the preparation

XX CC of coated core complexes for the invention.

XX SQ Sequence 40 AA:

Query Match 79.0%; Score 192; DB 22; Length 40;  
Best Local Similarity 90.0%; Pred. No. 1.1e-16;  
Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 GLFGAIAFGIENGWEGMIDGMYGKKKKKKKKKKKKKKKKKKKK 41  
Db 1 glfgaiegfiengwegmidgwygckkkkkkkkkkkkkkkkkkk 40

RESULT 3  
AAB13782  
ID AAB13782 standard; peptide: 24 AA.

XX AC AAB13782:  
XX DT 10-NOV-2000 (first entry)  
XX DE Soluble peptide antigen HA.  
XX KM HA peptide: cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;  
XX KM major histocompatibility complex class 1; MHC class 1; antigen; tumour;  
XX KM prostate; breast; multiple myeloma.  
XX OS Unidentified.  
XX PN WO200035949-A1.  
XX PD 22-JUN-2000.  
XX PF 14-DEC-1999; 99WO-US29724.  
XX PR 14-DEC-1998; 98US-0112324.  
XX PA (DEND-) DENDREON CORP.  
XX PI Laus R, Hakim I, Vidovic D;  
XX DR WPI: 2000-442365/38.  
XX PT Antigens modified by the covalent addition of a peptide that  
XX PT facilitates entry into antigen presenting cells, useful for producing  
XX PT compositions for immunizing against tumors and pathogens -  
XX PS Claim 2; Page 26; 34pp: English.

XX CC The present invention relates to compositions of modified soluble protein  
XX CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)  
XX CC response i.e. a major histocompatibility complex (MHC) class I molecule  
XX CC response. The protein antigen is modified by the covalent addition of a  
XX CC peptide sequence which facilitate entry of the antigen into antigen  
XX CC presenting cells (APCs). The present sequence is one such peptide  
XX CC sequence which can be used to modify the soluble antigens. The present  
XX CC sequence is peptide HA. The modified antigen composition may be used for  
XX CC immunising against, or treating a tumour e.g. prostate and breast  
XX CC carcinoma or multiple myeloma, or pathogen in mammals.

XX SQ Sequence 24 AA:

Query Match 58.8%; Score 143; DB 21; Length 24;  
Best Local Similarity 100.0%; Pred. No. 6.4e-11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAIAFGIENGWEGMIDGMYG 24  
Db 1 cglfgaiafgiengwegmidgwyg 24

```

RESULT ' 4
AAR26004
ID AAR26004 standard; peptide: 26 AA.
XX
AC AAR26004:
XX
DT 26-JAN-1993 (first entry)
XX
DE Influenza fusion peptide #1.
XX
KW Haemagglutinin; cysteine; liposome.
XX
OS Synthetic.
XX
PN EP497997-A.
XX
PD 12-AUG-1992.
XX
PF 02-FEB-1991: 91EP-0101414.
XX
PR 02-FEB-1991: 91EP-0101414.
XX
PA (NIKA-) NIKA HEALTH PROD LTD.
XX
PI Gluck R, Hermann P, Klein P;
XX
DR WPI: 1992-270078/33.
XX
PT Drug-contg. phospholipid bilayer vesicle with cell-specific
PT markers on membrane - where markers have at least 90 per cent
PT biological activity, used as pharmaceuticals against e.g. cancer
PT and AIDS
XX
PS Disclosure: Fig 2: 13pp; English.
XX
CC The sequences given in AAR26004-19 are influenza fusion peptides,
CC derived from the influenza virus haemagglutinin gene which were
CC used for the preparation of synthetic membrane vesicles. The
CC arrangement of at least one, pref. three cysteine residues at one
CC end of these peptides has been found useful for the fusion activity,
CC for the fusion of the liposome to the target cell membrane. The
CC liposomes produced using these peptides can contain at least one
CC active drug and can be used to target cells infected with cancer
CC or AIDS.
XX
SQ Sequence 26 AA:

Query Match 58.8%; Score 143; DB 13; Length 26;
Best Local Similarity 100.0%; Pred. No. 7e-11;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGLFGAIAAGFIENGWEGMIDGWG 24
   |||||||
DB 3 cglfgaiaagfiengwegmidgwyg 26

RESULT 5
AAM34269
ID AAM34269 standard; peptide: 26 AA.
XX
AC AAM34269:
XX
DT 14-MAY-1998 (first entry)
XX
DE Synthetic lipid vesicle fusion peptide 1.
XX
KW Virosome; vesicle; cationic lipid bilayer; viral fusion peptide;
KW drug delivery system; membrane; gene therapy; diagnosis; treatment;
KW cancer; leukemia; viral infection.
XX
OS Synthetic.
XX

```

```

PN W09741834-A1.
XX
PD 13-NOV-1997.
XX
PF 04-MAY-1997: 97WO-EP02268.
XX
PR 08-MAY-1996: 96EP-0107282.
XX
PA (NIKA-) NIKA HEALTH PROD LTD.
XX
PI Glueck R, Klein P, Maeltl ER;
XX
DR WPI: 1997-558673/51.
XX
PT Vesicle with cationic lipid bilayer that includes viral fusion
PT peptide - used for delivery of genetic material to cells, especially
PT for gene therapy of cancer, leukemia and viral infections
XX
PS Disclosure: Page 10; 52pp; English.
XX
CC Peptides AAM34269-W34284 represent novel lipid vesicles with positively
CC charged lipid bilayer membranes composed of a cationic and/or
CC polycationic lipid and at least one natural or synthetic viral fusion
CC peptide integrated in, or covalently linked to, the membrane. Such
CC peptides are used as drug delivery systems, preferably for
CC (non-)specific delivery of genetic material to target cells or tissues,
CC particularly for diagnosis, treatment (especially antisense treatment)
CC of cancer, leukemia and viral infections in humans or animals. Genetic
CC material is delivered, without infection, to resting or proliferating
CC cells, in vitro or in vivo. When the genetic material is an antisense
CC molecule, it is targeted to mRNA encoding a (proto)oncogene. The
CC continuous lipid layer does not leak. The peptides do not need to fuse
CC with, or destabilise, plasma membranes in order to enter the cytoplasm,
CC since the fusion peptide ensures cell penetration by endocytosis (after
CC which fusion of the vesicle and endosomal membrane occurs). The genetic
CC material thus has a greater chance of reaching the nucleus before it is
CC degraded or expelled. Transfer of the material is 1000-20000 times more
CC efficient than when using liposomes or conventional virosomes, so
CC smaller doses can be used, avoiding possible toxicity associated with the
CC genetic material.
XX
SQ Sequence 26 AA:

Query Match 58.8%; Score 143; DB 18; Length 26;
Best Local Similarity 100.0%; Pred. No. 7e-11;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGLFGAIAAGFIENGWEGMIDGWG 24
   |||||||
DB 3 cglfgaiaagfiengwegmidgwyg 26

RESULT 6
AAB70090
ID AAB70090 standard; peptide: 26 AA.
XX
AC AAB70090:
XX
DT 14-MAY-2001 (first entry)
XX
DE Cationic virosome crosslinker #1 for polypeptide binding.
XX
KW Cationic virosome; cytosolic; gene therapy; lipid bilayer vesicle;
KW viral haemagglutinin; drug delivery; cancer; leukemia; viral infection;
KW crosslinker.
XX
OS Unidentified.
XX
PN NZ504444-A.
XX
PD 24-NOV-2000.
XX

```



XX AAR98010-R98041 represent fusogenic peptides derived from spike  
CC glycoproteins of enveloped viruses. These fusogenic peptides act as  
CC endosome membrane disruption promoting components (EMDPCs). These  
CC sequences were used in a transfer moiety which makes up part of the  
CC multifunctional molecular complex of the invention. The multifunctional  
CC molecular complex is used for the transfer of a nucleic acid composition  
CC to a target cell, and comprises the nucleic acid composition and a  
CC transfer moiety comprising one or more cationic polyamine components  
CC bound to the nucleic acid composition. There are several different  
CC cationic polyamines that can be used in the complex, one of which  
CC contains one of these sequences linked via a bridging group to the main  
CC body of the cationic polyamine. The products can be used for delivering  
CC therapeutic agents, vaccines, foodstuffs, herbicides, plant growth  
CC regulators, insecticides, miticides, rodenticides, fungicides,  
CC parasiticides or nematocides. They can also be used for immunising an  
CC individual against a pathogen or for treating an autoimmune disease or  
CC hyperproliferating disease. The complexes provide for a high level of  
CC transfection and expression of the nucleic acid molecules in target  
CC cells.

XX Sequence 25 AA:

Query Match 55.1%; Score 134; DB 17; Length 25;  
Best Local Similarity 100.0%; Pred. No. 8.5e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMIDGMYG 24  
|||||  
Db 1 glfgaiagfiengwegmidgwyg 23

#### RESULT 9

AAR26005  
ID AAR26005 standard; peptide: 26 AA.

XX AAR26005;

AC 26-JAN-1993 (first entry)

DE Influenza fusion peptide #2.

KM Haemagglutinin; cystein; liposome.

OS Synthetic.

PN EP497997-A.

PD 12-AUG-1992.

PF 02-FEB-1991; 91EP-0101414.

PR 02-FEB-1991; 91EP-0101414.

PA (NIKA-) NIKA HEALTH PROD LTD.

PI Gluck R, Herrmann P, Klein P;

DR WPI; 1992-270078/33.

XX Drug-contg. phospholipid bilayer vesicle with cell-specific  
PT markers on membrane - where markers have at least 90 per cent  
PT biological activity, used as pharmaceuticals against e.g. cancer  
PT and AIDS

XX Disclosure; Fig 2; 13pp; English.

XX The sequences given in AAR26004-19 are influenza fusion peptides,  
CC derived from the influenza virus haemagglutinin gene which were  
CC used for the preparation of synthetic membrane vesicles. The  
CC arrangement of at least one, pref. three cystein residues at one  
CC end of these peptides has been found useful for the fusion activity,

CC for the fusion of the liposome to the target cell membrane. The  
CC liposomes produced using these peptides can contain at least one  
CC active drug and can be used to target cells infected with cancer  
CC or AIDS.

SO Sequence 26 AA:

Query Match 55.1%; Score 134; DB 13; Length 26;  
Best Local Similarity 100.0%; Pred. No. 8.8e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMIDGMYG 24  
|||||  
Db 1 glfgaiagfiengwegmidgwyg 23

#### RESULT 10

AAW34270  
ID AAW34270 standard; peptide: 26 AA.

AC AAW34270;

DT 14-MAY-1998 (first entry)

DE Synthetic lipid vesicle fusion peptide 2.

KM Virosome; vesicle; cationic lipid bilayer; viral fusion peptide;

KW drug delivery system; membrane; gene therapy; diagnosis; treatment;

KX cancer; leukaemia; viral infection.

OS Synthetic.

PN WO9741834-A1.

PD 13-NOV-1997.

PF 04-MAY-1997; 97WO-EP02268.

PR 08-MAY-1996; 96EP-0107282.

PA (NIKA-) NIKA HEALTH PROD LTD.

PI Glueck R, Klein P, Waelte ER;

DR WPI; 1997-558673/51.

XX Vesicle with cationic lipid bilayer that includes viral fusion  
PT peptide - used for delivery of genetic material to cells, especially  
PT for gene therapy of cancer, leukaemia and viral infections  
XX Disclosure; Page 10; 52pp; English.

XX Peptides AAW34269-W34284 represent novel lipid vesicles with positively  
CC charged lipid bilayer membranes composed of a cationic and/or  
CC polycationic lipid and at least one natural or synthetic viral fusion  
CC peptide integrated in, or covalently linked to, the membrane. Such  
CC peptides are used as drug delivery systems, preferably for  
CC (non-)specific delivery of genetic material to target cells or tissues,  
CC particularly for diagnosis, treatment (especially antisense treatment)  
CC of cancer, leukaemia and viral infections in humans or animals. Genetic  
CC material is delivered, without infection, to resting or proliferating  
CC cells, in vitro or in vivo. When the genetic material is an antisense  
CC molecule, it is targeted to mRNA encoding a (proto)oncogene. The  
CC continuous lipid layer does not leak. The peptides do not need to fuse  
CC with, or destabilise, plasma membranes in order to enter the cytoplasm,  
CC since the fusion peptide ensures cell penetration by endocytosis (after  
CC which fusion of the vesicle and endosomal membrane occurs). The genetic  
CC material thus has a greater chance of reaching the nucleus before it is  
CC degraded or expelled. Transfer of the material is 1000-20000 times more  
CC efficient than when using liposomes or conventional virosomes, so  
CC smaller doses can be used, avoiding possible toxicity associated with the  
CC genetic material.

XX Sequence 26 AA;  
SQ

Query Match 55.1%; Score 134; DB 18; Length 26;  
Best Local Similarity 100.0%; Pred. No. 8.8e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGMEGMDGMYG 24  
|||||  
Db 1 glfgaigfiengwegmidgwyg 23

RESULT 11  
AAB70091  
ID AAB70091 standard; peptide: 26 AA.  
XX  
AC AAB70091;  
XX  
DT 14-MAY-2001 (first entry)  
XX  
DE Cationic virosome crosslinker #2 for polypeptide binding.  
XX  
KM Cationic virosome; cytosstatic; gene therapy; lipid bilayer vesicle;  
KM viral haemagglutinin; drug delivery; cancer; leukaemia; viral infection;  
KM crosslinker.  
XX  
OS Unidentified.  
XX  
PN N2504444-A.  
XX  
PD 24-NOV-2000.  
XX  
PF 10-MAY-2000; 2000NZ-0504444.  
XX  
PR 10-MAY-2000; 2000NZ-0332666.  
XX  
PA (NIKA-) NIKA HEALTH PROD LTD.  
XX  
PI Walli ER, Gluck R, Klein P;  
XX  
DR WPI: 2001-233042/24.  
XX  
PT Lipid bilayer vesicle, useful for delivering drugs to target cells such  
PT as cancer, leukemic, or virally infected cells, comprises viral  
PT glycoproteins in positively charged membrane  
XX  
PS Disclosure: Page 14; 41pp; English.  
XX  
CC The present sequence is a crosslinker which is capable of linking to the  
CC surface of a novel cationic virosome and is capable of binding  
CC polypeptides. The cationic virosome is a lipid bilayer vesicle comprising  
CC a membrane with a net positive charge and contains 5-30 weight % based on  
CC total lipid, 1,3-dioleoyloxy-2-(6-carboxy-spermyl)-propyl-amine  
CC (DOSEPR), together with other lipids and at least one active fusogenic  
CC peptide. The fusogenic peptide is a viral haemagglutinin that causes the  
CC vesicle to be internalised by target cells through phagocytosis or  
CC endocytosis. The virosome is useful for delivering a desired drug or  
CC substance, preferably a nucleic acid, to target cells (resting or  
CC proliferating mammalian cells) such as cancer cells, leukemic cells or  
CC virally infected cells in vitro, in diagnostic or medical applications  
CC and for the manufacture of medicament for carrying out non-infectious  
CC gene therapy.  
XX  
SQ Sequence 26 AA;

Query Match 55.1%; Score 134; DB 22; Length 26;  
Best Local Similarity 100.0%; Pred. No. 8.8e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGMEGMDGMYG 24  
|||||  
Db 1 glfgaigfiengwegmidgwyg 23

Db 1 glfgaigfiengwegmidgwyg 23

RESULT 12  
AAR98024  
ID AAR98024 standard; peptide: 28 AA.  
XX  
AC AAR98024;  
XX  
DT 12-NOV-1996 (first entry)  
XX  
DE Fusogenic peptide derived from HA-2 X:31 spike glycoprotein.  
XX  
KM Fusogenic peptide; spike glycoprotein; enveloped virus; vaccine; EMDPC;  
KM endosome membrane disruption promoting component; cationic polyamine;  
KM multifunctional molecular complex; foodstuff; herbicide; insecticide;  
KM plant growth regulator; miticide; rodenticide; fungicide; parasiticide;  
KM nematocide; immunisation; pathogen; therapy; autoimmune disease;  
KM hyperproliferating disease.  
XX  
OS Synthetic.  
XX  
PN W09610038-A1.  
XX  
PD 04-APR-1996.  
XX  
PF 28-SEP-1995; 95WO-US12502.  
XX  
PR 28-SEP-1994; 94US-0314060.  
XX  
PA (APOL-) APOLLON INC.  
XX  
PI Boutin RH;  
XX  
DR WPI: 1996-200887/20.  
XX  
PT New complexes for nucleic acid transfer to target cells - comprising  
PT a nucleic acid compsn. and a cationic polyamine with an endosome  
PT membrane disruption component.  
XX  
PS Disclosure: Page 30; 138pp; English.  
XX  
CC AAR98010-R98041 represent fusogenic peptides derived from spike  
CC glycoproteins of enveloped viruses. These fusogenic peptides act as  
CC endosome membrane disruption promoting components (EMDPCs). These  
CC sequences were used in a transfer moiety which makes up part of the  
CC multifunctional molecular complex of the invention. The multifunctional  
CC molecular complex is used for the transfer of a nucleic acid composition  
CC to a target cell, and comprises the nucleic acid composition and a  
CC transfer moiety comprising one or more cationic polyamine components  
CC bound to the nucleic acid composition. There are several different  
CC cationic polyamines that can be used in the complex, one of which  
CC contains one of these sequences linked via a bridging group to the main  
CC body of the cationic polyamine. The products can be used for delivering  
CC therapeutic agents, vaccines, foodstuffs, herbicides, plant growth  
CC regulators, insecticides, miticides, rodenticides, fungicides,  
CC parasiticides or nematocides. They can also be used for immunising an  
CC individual against a pathogen or for treating an autoimmune disease or  
CC hyperproliferating disease. The complexes provide for a high level of  
CC transfection and expression of the nucleic acid molecules in target  
CC cells.  
XX  
SQ Sequence 28 AA;

Query Match 55.1%; Score 134; DB 17; Length 28;  
Best Local Similarity 100.0%; Pred. No. 9.5e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGMEGMDGMYG 24  
|||||  
Db 1 glfgaigfiengwegmidgwyg 23





```

XX OS Virididae.
XX PN WO200164013-A2.
XX PD 07-SEP-2001.
XX PF 07-FEB-2001; 2001WO-US03988.
XX PR 29-FEB-2000; 2000US-0515965.
XX PA (TRIM-) TRIMERIS INC.
XX PI Antczak JB, Delmedico MK, Erickson JB, Lambert DM, Sista P;
XX DR WPI: 2001-514829/56.
XX PT Heptad repeat region peptide analogs useful for inhibiting virus/cells
XX PT fusion, useful for treating HIV and Respiratory Syncytial Virus
XX PT Infection.
XX PS Disclosure: Page 448; 587pp; English.
XX CC The invention relates to isolated analogues of the heptad repeat region
XX CC peptides DP178 and DP107. DP178 and DP107 correspond to amino acids
XX CC 638-673 (heptad repeat region HR2) and 558-595 (heptad repeat region
XX CC HR1) respectively, of HIV-1LAI transmembrane protein gp41. The HR1
XX CC and HR2 regions of proteins interact non-covalently with each other
XX CC and/or with peptides derived from them. This interaction is required for
XX CC normal infectivity of viruses such as RSV and HIV. The heptad
XX CC repeat region peptide analogues may be used to inhibit respiratory
XX CC syncytial virus (RSV) infection in a cell. They may also be used to
XX CC inhibit HIV infection. The present sequence is a peptide provided in
XX CC the specification.
SQ Sequence 30 AA:

Query Match 55.1%; Score 134; DB 22; Length 30;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIFAGFIENGEGMIDGMYG 24
Db 1 gltgalaigfiengwegmldgwyg 23

```

Search completed: July 1, 2002, 06:19:25  
 Job time: 512 sec

! FINDPATTERNS on geneseqp:\* allowing 0 mismatches

! 1 C(R,K){20} July 1, 2002 07:07 ..

1 AAW45801 ck: 9500 len: 39 ! Aaw45801 One chain of a bombesin dimer. 6/1

C(R,K){20}  
C(K){20}  
1: CXXXXXXXXXXXXXXXXX XGCGQ

1 AAB13780 ck: 7317 len: 21 ! Aab13780 Soluble peptide antigen PK. 11/200

C(R,K){20}  
C(K){20}  
1: CXXXXXXXXXXXXXXXXX

1 AAU18238 ck: 5509 len: 58 ! Aau18238 Novel human DNA-binding protein #8

C(R,K){20}  
C(K){20}  
35: KFYFV CXXXXXXXXXXXXXXXXX KKK

1 AAO03766 ck: 8808 len: 81 ! Aao03766 Human polypeptide SEQ ID NO 17658.

C(R,K){20}  
C(K){20}  
48: LTTTA CXXXXXXXXXXXXXXXXX KKKKK

1 AAO11210 ck: 863 len: 70 ! Aao11210 Human polypeptide SEQ ID NO 25102.

C(R,K){20}  
C(K){20}  
30: IDLCL CXXXXXXXXXXXXXXXXX KKKKK

Databases searched:

EMBL, Release 7.0, Released on 28Apr2002, Formatted on 2May2002

Total finds: 5  
Total length: 114,001,827  
Total sequences: 766,495  
CPU time: 07:03.10

Seq #6

!!AA\_SEQUENCE 1.0  
ID AAW45801 standard; peptide; 39 AA.  
XX  
AC AAW45801:  
XX  
DT 25-JUN-1998 (first entry)  
XX  
DE One chain of a bombesin dimer.  
XX  
KW Alpha-melanocyte stimulating hormone; alpha-MSH; receptor agonist;  
KW alpha-MSH-Ant; bombesin; dimer; bivalent agonist; disulphide bond;  
KW G-protein coupled receptor.  
XX  
OS Synthetic.  
XX  
FT Key Location/Qualifiers  
FT Cross-links 1  
FT /note- "This residue is disulphide bonded to the  
FT corresponding Cys residue of an identical chain  
FT to form a dimer."  
FT Modified-site 28  
FT /note- "Epsilon-aminohexanoic acid"  
XX  
PN WO9803632-A1.  
XX  
PD 29-JAN-1998.  
XX  
PF 23-JUL-1997; 97WO-US12911.  
XX  
PR 24-JUL-1996; 96US-0686934.  
XX  
PA (UYVA ) UNIV YALE.  
XX  
PI Carltthers MD, Lerner MR:  
XX  
DR WPI: 1998-120757/11.  
XX  
XX Bivalent agonist of G-protein coupled receptors containing two  
XX ligand domains - bonded to molecular backbone, for treatment of  
XX hypertension, promotion of skin tanning etc., also for delivering  
XX drugs and gene therapy vectors to selected cells  
XX  
PS Claim 41: Page 48; 71pp; English.  
XX  
CC This sequence represents one of two identical chains disulphide bonded  
CC to form a bombesin dimer. The invention relates to bivalent agonists,  
CC with affinity for at least 1 G-protein coupled receptor (GPCR). The  
CC bivalent agonists comprise: (a) two ligand domains (LD), individually  
CC agonists or antagonists for GPCR, spaced 40-250 Angstrom apart, and (b)  
CC a molecular backbone (MB) covalently bound to LD. The bivalent agonists  
CC are useful in human or veterinary medicine as carriers for drugs or gene  
CC therapy vectors, allowing these to be endocytosed by GPCR-expressing  
CC cells. They can also be used e.g. to treat hypertension (angiotensin-  
CC based LD); to increase levels of luteinising hormone (LH), using LD  
CC derived from LH-releasing hormone, or to promote skin tanning (LD based  
CC on alpha -melanocyte-stimulating hormone, MSH). The bivalent agonists  
CC are administered orally, by injection or topically. Typical doses for  
CC skin tanning are 1-4000 (especially 30-100) mu mole/kg systemically or  
CC the bivalent agonists are administered topically in a composition  
CC containing 0.001-10 (especially 1) mM. Where both LD are agonists, the  
CC bivalent agonist has a synergistically higher activity than two  
CC individual agonist ligands, and where at least 1 is an antagonist the  
CC effect is stimulatory. The bivalent agonists are active at lower  
CC concentrations than known agonists so should avoid toxicity problems.  
XX  
SQ Sequence 39 AA;  
AAW45801 Length: 39 July 1, 2002 14:50 Type: P Check: 9500  
1 CKKKKKKKKK KKKKKKKKK KXGGGQORLG NOMAVGHLM  
!!AA\_SEQUENCE 1.0  
ID AAB13780 standard; peptide; 21 AA.

XX  
AC AAB13780;  
XX  
DT 10-NOV-2000 (first entry)  
XX  
DE Soluble peptide antigen pK.  
XX  
KW pK peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;  
KW major histocompatibility complex class I; MHC class I; antigen; tumour;  
KW prostate; breast; multiple myeloma.  
XX  
OS Unidentified.  
XX  
PN WO200035949-A1.  
XX  
PD 22-JUN-2000.  
XX  
PF 14-DEC-1999; 99WO-US29724.  
XX  
PR 14-DEC-1998; 98US-0112324.  
XX  
PA (DEND-) DENDREON CORP.  
XX  
PI Laus R, Hakim I, Vidovic D;  
XX  
DR WPI: 2000-442365/38.  
XX  
XX Antigens modified by the covalent addition of a peptide that  
XX facilitates entry into antigen presenting cells, useful for producing  
XX compositions for immunizing against tumors and pathogens -  
XX  
PS Claim 2: Page 26; 34pp; English.  
XX  
CC The present invention relates to compositions of modified soluble protein  
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)  
CC response i.e. a major histocompatibility complex (MHC) class I molecule  
CC response. The protein antigen is modified by the covalent addition of a  
CC peptide sequence which facilitate entry of the antigen into antigen  
CC presenting cells (APCs). The present sequence is one such peptide  
CC sequence which can be used to modify the soluble antigens. The present  
CC sequence is peptide pK. The modified antigen composition may be used for  
CC immunising against, or treating a tumour e.g. prostate and breast  
CC carcinoma or multiple myeloma, or pathogen in mammals.  
XX  
SQ Sequence 21 AA;  
AAB13780 Length: 21 July 1, 2002 14:50 Type: P Check: 7317  
1 CKKKKKKKKK KKKKKKKKK K  
!!AA\_SEQUENCE 1.0  
ID AAU18238 standard; Protein; 58 AA.  
XX  
AC AAU18238;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Novel human DNA-binding protein #85.  
XX  
KW Human: DNA-binding protein; histone; chromo domain protein;  
KW chromatin organisation modifier; Y-box binding protein;  
KW DNA organisation; gene transcription; malignant disease;  
KW autoimmune disorder; rheumatic disease; genetic abnormality;  
KW infectious disease; neurological disorder; gene therapy;  
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;  
KW cytostatic.  
XX  
OS Homo sapiens.  
XX  
PN WO20015162-A1.  
XX  
PD 02-AUG-2001.



PT Nucleic acid molecules encoding human secreted chromosomal binding  
 PT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
 PT Alzheimer's and Parkinson's diseases and cancers -  
 PS Claim 11: SEQ ID NO 223: 561pp: English.

CC The present invention relates to the isolation of novel DNA-binding  
 CC proteins, and cDNA (AA529030-AMS29157) and genomic sequences encoding  
 CC for these proteins. DNA-binding proteins such as histones, chromo  
 CC (chromatin organisation modifier) domain proteins, and Y-box binding  
 CC proteins may contribute to diseases resulting from aberrant DNA  
 CC organisation and/or gene transcription. The sequences of the invention  
 CC are useful in screening assays to identify antagonists and/or agonists  
 CC that may enhance or block activities mediated by DNA-binding proteins.  
 CC Blockers of DNA-binding proteins may be useful in treating disorders  
 CC such as malignant diseases (e.g. cancer), autoimmune disorders  
 CC (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid  
 CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious  
 CC diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's  
 CC disease). The polynucleotide sequences of the invention may also be  
 CC used in gene therapy. AAU18154-AAU18281 represent novel DNA-binding  
 CC proteins.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 58 AA:

AAU18238 Length: 58 July 1, 2002 14:50 Type: P Check: 5509 ..

1 TYLECEHNSL VNSKCLTVVL SRCISVCLNK FYFVCKKKKK KKKKKKKKK

51 KKKKKKKK

!!AA\_SEQUENCE 1.0  
 ID AA003766 standard: Protein: 81 AA.

AC AA003766;

XX 06-NOV-2001 (first entry)

DT Human polypeptide SEQ ID NO 17658.

DE Human polypeptide SEQ ID NO 17658.

XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

OS WO200164835-A2.

PN 07-SEP-2001.

PD 26-FEB-2001: 2001WO-US04927.

PF 28-FEB-2000: 2000US-0515126.

PR 18-MAY-2000: 2000US-0577409.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT;

PI WPI: 2001-514838/56.

DR N-PSDB: AA183697.

XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -

PS Claim 20: SEQ ID NO 17658; 1399pp + Sequence Listing: English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and

CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 81 AA:

AA003766 Length: 81 July 1, 2002 14:50 Type: P Check: 8808 ..

1 GLNQTLRKRI LAYSSITHIG XIIVLPYP NITLNLTIY ILLTTACKK

51 KKKKKKKKK KKKKKKKKK KKKKKGGCG A

ID AA011210 standard: Protein: 70 AA.

AC AA011210;

XX 06-NOV-2001 (first entry)

DT Human polypeptide SEQ ID NO 25102.

DE Human polypeptide SEQ ID NO 25102.

XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

OS WO200164835-A2.

PN 07-SEP-2001.

PD 26-FEB-2001: 2001WO-US04927.

PF 28-FEB-2000: 2000US-0515126.

PR 18-MAY-2000: 2000US-0577409.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT;

PI WPI: 2001-514838/56.

DR N-PSDB: AA191141.

XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -

PS Claim 20: SEQ ID NO 25102; 1399pp + Sequence Listing: English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and

CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ . Sequence 70 AA;

AA011210 Length: 70 July 1, 2002 14:50 Type: P Check: 863 ..

1 YIHRTTFM CMNXGLKDNV DKXTIDLCLC KKKKKKKKK KKKKKKKKK

51 KKKKKKKKK KKKKPGGGA

---

!!SEQUENCE\_LIST 1.0  
! FINDPATTERNS on geneseqp: \* allowing 0 mismatches

! 1 C(R,K){20}

July 1, 2002 14:38 ..

GENESEQP1990S:AAW45801 ck: 9500 len: 39 finds: 1 ! Aaw45801 One chain of a bomb  
GENESEQP2000:ABD13780 ck: 7317 len: 21 finds: 1 ! Abd13780 Soluble peptide anti  
GENESEQP2001:AAU18238 ck: 5509 len: 58 finds: 1 ! Aau18238 Novel human DNA-bind  
GENESEQP2001:AAO03766 ck: 8808 len: 81 finds: 1 ! Aao03766 Human polypeptide SE  
GENESEQP2001:AAO11210 ck: 863 len: 70 finds: 1 ! Aao11210 Human polypeptide SE

\\End of list

Databases searched:

EMBL, Release 7.0, Released on 28Apr2002, Formatted on 2May2002

Total finds: 5  
Total length: 114,001,827  
Total sequences: 766,495  
CPU time: 09:53.67

1 FINDPATTERNS on pir:\* allowing 0 mismatches

1 C(R,K){20} July 1, 2002 06:14 ..

Databases searched:

MBRF, Release 71.0, Released on 15Feb2002, Formatted on 19Feb2002

Total finds: 0  
Total length: 96,089,334  
Total sequences: 283,138  
CPU time: 04:30.92



! FINDPATTERNS on swp:\* allowing 0 mismatches  
! 1 C(R,K){20}

July 1, 2002 06:15 ..

Databases searched:

SWISS-PROT, Release 40.1, Released on 11Feb2002, Formatted on 19Feb2002  
SPRMBL, Release 19.0, Released on 14Dec2001, Formatted on 19Feb2002

Total finds: 0  
Total length: 211,714,479  
Total sequences: 667,446  
CPU time: 10:17.12

! FINDPATTERNS on pir:\* allowing 0 mismatches

1 (R,K){20}

July 1, 2002 06:48 ..

1 T49173 ck: 4143 len: 517 1 hypothetical protein T20N10.250 - Arabidops

(R,K){20}

444: FEHVG KKKKKKKKKKKKKKKKKKK KKKIR

(K){20}

445: ERVGG KKKKKKKKKKKKKKKKKKK KKTIRL

(K){20}

446: RVGKK KKKKKKKKKKKKKKKKKKK KIRLN

(K){20}

447: VGKKK KKKKKKKKKKKKKKKKKKK IRLNF

1 S58321 ck: 1384 len: 126 1 probable membrane protein YOR309c - yeast

(R,K){20}

53: RKRRT RRRRRRRRRRRRRRRRRR KRSPR

(R,K){20}

54: KRRTT RRRRRRRRRRRRRRRRRR RSPRK

(R,K){20}

55: RRTTR RRRRRRRRRRRRRRRR SPRKR

1 T46395 ck: 7330 len: 380 1 hypothetical protein DKF2p434I1120.1 - huma

(R,K){20}

355: NULLQ KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

356: LLLQK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

357: LLOKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

358: LQKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

359: OKKKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

360: KKKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

361: KKKKK KKKKKKKKKKKKKKKKKKK

1 I52523 ck: 8048 len: 215 1 nucleoporin p62 homolog - rat (fragment)

(R,K){20}

35: CEFLK KKKKKKKKKKKKKKKKKKK KKTGD

(K){20}

36: EFLEK KKKKKKKKKKKKKKKKKKK KTGDN

(K){20}

37: FLEKK KKKKKKKKKKKKKKKKKKK TGDNA

Databases searched:

NBRF, Release 71.0, Released on 15Feb2002, Formatted on 19Feb2002  
Total finds: 17  
Total length: 96,089,334  
Total sequences: 283,138  
CPU time: 09:02.03

Seq # 7

!!SEQUENCE LIST 1 0  
! FINDPATTERNS on pir:\* allowing 0 mismatches

1 (R,K){20}

July 1, 2002 14:57

PIR2:T49173 ck: 4143 len: 517 finds: 4 ! hypothetical protein T20N10.25  
PIR2:S58321 ck: 1384 len: 126 finds: 3 ! probable membrane protein YOR3  
PIR2:T46395 ck: 7330 len: 380 finds: 7 ! hypothetical protein DKFZp434J  
PIR2:I52523 ck: 8048 len: 215 finds: 3 ! nucleoporin p62 homolog - rat  
\\End of list

Databases searched:

NBRF, Release 71.0, Released on 15Feb2002, Formatted on 19Feb2002

Total finds: 17  
Total length: 96,089,334  
Total sequences: 283,138  
CPU time: 08:41.42

!!AA\_SEQUENCE 1.0  
PI:T49173 - hypothetical protein T20N10.250 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 08-Dec-2000  
C:Accession: T49173  
R:D'Angelio, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.;  
Rudd, S.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat, M.  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: 225017  
A:Accession: T49173  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-517 <DNA>  
A:Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250  
A:Experimental source: cultivar Columbia; BAC clone T20N10  
C:Genetics:  
A:Gene: ATSP:T20N10.250  
A:Map position: 3  
A:Introns: 312/3; 359/3; 444/3  
C:Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30  
T49173 Length: 517 July 1, 2002 15:06 Type: P Check: 4143 ..  
1 MDLFSSLPNE LLYHLSFLS TKEALTSVL SKRMRLFAF VPYLEFDSV  
51 FLHPEERKE KEGIQSFMD FVDRVLDLHG DSLIKTFSLK CKTGVDSDHV  
101 DMWICNVLAR GVSDDLDFID FRDIYSLPHE VGVSRLLVYL RVGSSEDLVW  
151 WQKFLCLPML KTVLVDSCL CIGQFQILL ACPALELDM TWTRKSDNV  
201 TVSSSLKEL TIDLHGCCSV VNLKLSFDA PSLVYFYDCC SLAEDEYQVN  
251 LKNLVEAOIN LLLTQAOIEQ VRALNENMLV ADDVFGGLN ANKLITGLRN  
301 VQQLYLSPT LEVSRCECG MPVFNNLKV L SIMSDNRGW QAMPYLLRNC  
351 PLETLIEG LLHYATDKCG DVCDCISRDY KDHSLTSCP KKLQIYEFRC  
401 TIRELEMIKH FLKIFPCLE MDIYAHENSH TLFKDPITFE RVGKKKKKKK  
451 KKKKKKKKK KKKKKKIRLN FKPVNKTQF LKRLADKLCF IPQCLEFLDV  
501 DSSLGELALL AMDSRPS  
!!AA\_SEQUENCE 1.0  
PI:S58321 - probable membrane protein YOR309c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein O6105  
C:Species: Saccharomyces cerevisiae  
C:Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 21-Jul-2000  
C:Accession: S58321; S67215; S71989  
R:Pearson, B.M.; Hernandez, Y.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: S58318  
A:Accession: S58321  
A:Molecule type: DNA  
A:Residues: 1-126 <PEP>  
A:Cross-references: EMBL:X90565; NID:9440836; PID:9440840  
R:Pearson, B.M.; Hernandez, Y.; Kalogeropoulos, A.; Schweizer, M.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67213  
A:Accession: S67215  
A:Molecule type: DNA  
A:Residues: 1-126 <PEP>  
A:Cross-references: EMBL:Z75217; NID:91420680; PID:e252431; PID:91420681;  
MIPS:YOR309C  
A:Experimental source: strain S288C  
R:Pearson, B.M.; Hernandez, Y.; Payne, J.; Wolf, S.S.; Kalogeropoulos, A.;  
Schweizer, M.  
yeast 12, 1021-1031, 1996  
A:Title: Sequencing of a 35.71 kb DNA segment on the right arm of yeast  
chromosome XV reveals regions of similarity to chromosomes I and XIII.  
A:Reference number: S71986; MUID:97051589

A:Accession: S71989  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-126 <PEP>  
A:Cross-references: EMBL:X90565; NID:9440836; PID:CA62164.1; PID:9440840  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August  
1995  
C:Genetics:  
A:Map position: 15R  
C:Keywords: transmembrane protein  
F:3-19/Domain: transmembrane #status predicted <TM1>  
F:107-123/Domain: transmembrane #status predicted <TM2>

S58321 Length: 126 July 1, 2002 15:06 Type: P Check: 1384 ..

1 MGLIIFQRL LILNPIIMWK RKKRKKRRK RERETMMKIP RILKRLRRR  
51 RTRRRKKRR KRRKKRRRR RKRSPKRRR KRRNKAFYI LIISDPSRL  
101 LFGFRKFSII IQCLTFSPFH ILFHNH  
!!AA\_SEQUENCE 1.0  
PI:T46395 - hypothetical protein DKFZp43411120.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T46395  
R:Ostenwelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: 223031  
A:Accession: T46395  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-380 <AAA>  
A:Cross-references: EMBL:AL137556  
A:Experimental source: adult testis; clone DKFZp43411120  
C:Genetics:  
A>Note: DKFZp43411120.1

T46395 Length: 380 July 1, 2002 15:06 Type: P Check: 7330 ..

1 MGSTDSKLNK RKAVIQLTFTK TQPEATDGA FMOGFNADTA TSVQDVFALV  
51 PAAETRAVRE ESPSNLATLC YKAVEKLYOG AESGCHSEKE KOIYLCNSRL  
101 LTRVLPIEFE DPMWRGFFWS TVPGAGRGCG EEDDEHARPL AESLLAIAID  
151 LIFCPDFTVQ SHRSTVSDA EDVHSLDSC YIWEAGYGFA HSPQPVYIHD  
201 MRMELKLL LTCRSEAMYL PPAPEGSTN PMVQFCSTE NRRHALPLFTS  
251 LINTVCAYDP VGYGIPYNH LFSDYREPLY EEAQVLIYV LDHDSASSAS  
301 PIVDGTITCT AMDADPPGP ENLFVNYLSR IHREDFOPFI LKGIRALLSN  
351 LLLQKKKKKK KKKKKKKKK KKKKKKKKK

!!AA\_SEQUENCE 1.0  
PI:I52523 - nucleoporin p62 homolog - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 27-Feb-1997 #sequence\_revision 27-Feb-1997 #text\_change 05-Nov-1999  
C:Accession: I52523  
R:Wang, Z.Q.; Akmal, K.M.; Kim, K.H.  
Biol. Reprod. 51, 1022-1030, 1994  
A:Title: An unusual nucleoporin-related messenger ribonucleic acid is present  
in the germ cells of rat testis.  
A:Reference number: I52523; MUID:95151924  
A:Accession: I52523  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-215 <RES>  
A:Cross-references: GB:S75997; NID:9913245; PIDN:AB3384.1; PID:9913246  
A:Experimental source: testis

152523 Length: 215 July 1, 2002 15:06 Type: P Check: 8048 ..  
1 SGRATSSCD EDCLSSLPF SLSGPVKODC EFLEKKKKKK KKKKKKKKK  
51 KKKKKKTGDN AKSVSRQYSL KVKLEHEAE QAKVELDFIL SQKELEDLL  
101 SPLEESVKEQ SGTIYLQHAD EEREKTYKLA ENIDAQLKRW AQLKDIIEH  
151 LNMAGGPADT SDPLQIQICKI LNAHMSLQW VDQSSALLQR RVEEASRVCE  
201 SRKREQERSL RIAFD

---

```
i FINDPATTERNS on swp:* allowing 0 mismatches
i
1 (R,K){20}
      July 1, 2002 06:38 ..
      Q12444 ck: 1384 len: 126 i Q12444 saccharomyces cerevisiae (baker's ye
1      (R,K){20}
      (R,K){20}
53: KKRRT KKKKKKKKKKKKKKKKKRRR KKSPR
      (R,K){20}
54: KRRTT KKKKKKKKKKKKKKKKKRRR RSPRK
      (R,K){20}
55: RRTTR KKKKKKKKKKKKKKKKKRRR SPRKR
      Q9P529 ck: 291 len: 128 i Q9P529 neurospora crassa. hypothetical 15.2
1      (R,K){20}
      (K){20}
71: KKKKQ KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
72: KKNQK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
73: KNQKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
74: NQKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
75: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
76: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
77: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
78: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
79: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
80: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
81: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
82: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      Q9HC48 ck: 7602 len: 667 i Q9HC48 homo sapiens (human). ctcl tumor ant
1      (R,K){20}
      (R,K){20}
648: GDKTD KKKKKKKKKKKKKKKKKKKK
      Q9H6Q7 ck: 3351 len: 720 i Q9H6Q7 homo sapiens (human). cdna: flj21979
1      (R,K){20}
      (K){20}
692: IVSIS KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
1      (K){20}
693: VSISK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
694: SISKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
695: ISKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
696: SKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
697: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
698: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
699: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
700: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
701: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      Q9H5V6 ck: 379 len: 168 i Q9H5V6 homo sapiens (human). cdna: flj22
1      (R,K){20}
      (K){20}
140: VREWE KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
141: REWEK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
142: EWEKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
143: WEKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
144: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
145: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
146: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
147: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
148: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
149: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      Q9NT34 ck: 7330 len: 380 i Q9NT34 homo sapiens (human). hypotheticala
1      (R,K){20}
      (K){20}
355: NLLIQ KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
356: LLLQK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
```

357: LLOKK KKKKKKKKKKKKKKKKK KKKK  
          (K){20}  
358: LOKKK KKKKKKKKKKKKKKKKK KKK  
          (K){20}  
359: OKKKK KKKKKKKKKKKKKKKKK KK  
          (K){20}  
360: KKKKK KKKKKKKKKKKKKKKKK K  
          (K){20}  
361: KKKKK KKKKKKKKKKKKKKKKK

Q95LV6 ck: 7515 len: 531 ! Q95LV6 macaca fascicularis (crab eating mac  
          (R,K){20}  
          (K){20}

502: YKGNK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
503: KGNKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
504: GNSKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
505: NSKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
506: SKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
507: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
508: KKKKK KKKKKKKKKKKKKKKKK KKKK  
          (K){20}  
509: KKKKK KKKKKKKKKKKKKKKKK KKK  
          (K){20}  
510: KKKKK KKKKKKKKKKKKKKKKK KK  
          (K){20}  
511: KKKKK KKKKKKKKKKKKKKKKK K  
          (K){20}  
512: KKKKK KKKKKKKKKKKKKKKKK

Q9LXR2 ck: 4143 len: 517 ! Q9LXR2 arabidopsis thaliana (mouse-ear cress  
          (R,K){20}  
          (K){20}

444: FERYG KKKKKKKKKKKKKKKKK KKKIR  
          (K){20}  
445: ERYGK KKKKKKKKKKKKKKKKK KIRL  
          (K){20}  
446: RVGKK KKKKKKKKKKKKKKKKK KIRLN  
          (K){20}  
447: VGRKK KKKKKKKKKKKKKKKKK IRLNF  
          (K){20}

Q9LGZ9 ck: 6094 len: 260 ! Q9LGZ9 arabidopsis thaliana (mouse-ear cress  
          (R,K){20}  
          (K){20}  
6: MDRCI KKKKKKKKKKKKKKKKK KKKKK

7: DRCIR KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
8: RCIRK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
9: CIRKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
10: IRKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
11: RKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
12: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
13: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
14: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
15: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
16: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
17: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
18: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
19: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
20: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
21: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
22: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
23: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
24: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
25: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
26: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
27: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
28: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
29: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
          (K){20}  
30: KKKKK KKKKKKKKKKKKKKKKK KKKKK

31: kkkk (k){20} kkkk  
32: kkkk (k){20} kkkk  
33: kkkk (k){20} kkkk  
34: kkkk (k){20} kkkk  
35: kkkk (k){20} kkkk  
36: kkkk (k){20} kkkk  
37: kkkk (k){20} kkkk  
38: kkkk (k){20} kkkk  
39: kkkk (k){20} kkkk  
40: kkkk (k){20} kkkk  
41: kkkk (k){20} kkkk  
42: kkkk (k){20} kkkk  
43: kkkk (k){20} kkkk  
44: kkkk (k){20} kkkk  
45: kkkk (k){20} kkkk  
46: kkkk (k){20} kkkk  
47: kkkk (k){20} kkkk  
48: kkkk (k){20} kkkk  
49: kkkk (k){20} kkkk  
50: kkkk (k){20} kkkk  
51: kkkk (k){20} kkkk  
52: kkkk (k){20} kkkk  
53: kkkk (k){20} kkkk  
54: kkkk (k){20} kkkk

-----

55: kkkk kkkkkkkkkkkkkkkkk kkkk  
56: kkkk (k){20} kkkk  
57: kkkk (k){20} kkkk  
58: kkkk (k){20} kkkk  
59: kkkk (k){20} kkkk  
60: kkkk (k){20} kkkk  
61: kkkk (k){20} kkkk  
62: kkkk (k){20} kkkk  
63: kkkk (k){20} kkkk  
64: kkkk (k){20} kkkk  
65: kkkk (k){20} kkkk  
66: kkkk (k){20} kkkk  
67: kkkk (k){20} kkkk  
68: kkkk (k){20} kkkk  
69: kkkk (k){20} kkkk  
70: kkkk (k){20} kkkk  
71: kkkk (k){20} kkkk  
72: kkkk (k){20} kkkk  
73: kkkk (k){20} kkkk  
74: kkkk (k){20} kkkk  
75: kkkk (k){20} kkkk  
76: kkkk (k){20} kkkk  
77: kkkk (k){20} kkkk  
78: kkkk (k){20} kkkk  
79: kkkk (k){20} kkkk



80: kkkk (k){20} kkkk  
81: kkkk (k){20} kkkk  
82: kkkk (k){20} kkkk  
83: kkkk (k){20} kkkk  
84: kkkk (k){20} kkkk  
85: kkkk (k){20} kkkk  
86: kkkk (k){20} kkkk  
87: kkkk (k){20} kkkk  
88: kkkk (k){20} kkkk  
89: kkkk (k){20} kkkk  
90: kkkk (k){20} kkkk  
91: kkkk (k){20} kkkk  
92: kkkk (k){20} kkkk  
93: kkkk (k){20} kkkk  
94: kkkk (k){20} kkkk  
95: kkkk (k){20} kkkk  
96: kkkk (k){20} kkkk  
97: kkkk (k){20} kkkk  
98: kkkk (k){20} kkkk  
99: kkkk (k){20} kkkk  
100: kkkk (k){20} kkkk  
101: kkkk (k){20} kkkk  
102: kkkk (k){20} kkkk  
103: kkkk (k){20} kkkk

104: kkkk (k){20} kkkk  
105: kkkk (k){20} kkkk  
106: kkkk (k){20} kkkk  
107: kkkk (k){20} kkkk  
108: kkkk (k){20} kkkk  
109: kkkk (k){20} kkkk  
110: kkkk (k){20} kkkk  
111: kkkk (k){20} kkkk  
112: kkkk (k){20} kkkk  
113: kkkk (k){20} kkkk  
114: kkkk (k){20} kkkk  
115: kkkk (k){20} kkkk  
116: kkkk (k){20} kkkk  
117: kkkk (k){20} kkkk  
118: kkkk (k){20} kkkk  
119: kkkk (k){20} kkkk  
120: kkkk (k){20} kkkk  
121: kkkk (k){20} kkkk  
122: kkkk (k){20} kkkk  
123: kkkk (k){20} kkkk  
124: kkkk (k){20} kkkk  
125: kkkk (k){20} kkkk  
126: kkkk (k){20} kkkk  
127: kkkk (k){20} kkkk

128: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
129: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
130: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
131: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
132: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
133: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
134: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
135: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
136: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
137: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
138: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
139: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
140: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
141: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
142: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
143: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
144: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
145: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
146: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
147: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
148: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
149: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
150: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
151: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
152: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}

153: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
154: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
155: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
156: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
157: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
158: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
159: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
160: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
161: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
162: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
163: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
164: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
165: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
166: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
167: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
168: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
169: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
170: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
171: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
172: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
173: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
174: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
175: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
176: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}

177: kkkk (K)120  
178: kkkk (K)120  
179: kkkk (K)120  
180: kkkk (K)120  
181: kkkk (K)120  
182: kkkk (K)120  
183: kkkk (K)120  
184: kkkk (K)120  
185: kkkk (K)120  
186: kkkk (K)120  
187: kkkk (K)120  
188: kkkk (K)120  
189: kkkk (K)120  
190: kkkk (K)120  
191: kkkk (K)120  
192: kkkk (K)120  
193: kkkk (K)120  
194: kkkk (K)120  
195: kkkk (K)120  
196: kkkk (K)120  
197: kkkk (K)120  
198: kkkk (K)120  
199: kkkk (K)120  
200: kkkk (K)120

201: kkkk  
202: kkkk (K)120  
203: kkkk (K)120  
204: kkkk (K)120  
205: kkkk (K)120  
206: kkkk (K)120  
207: kkkk (K)120  
208: kkkk (K)120  
209: kkkk (K)120  
210: kkkk (K)120  
211: kkkk (K)120  
212: kkkk (K)120  
213: kkkk (K)120  
214: kkkk (K)120  
215: kkkk (K)120  
216: kkkk (K)120  
217: kkkk (K)120  
218: kkkk (K)120  
219: kkkk (K)120  
220: kkkk (K)120  
221: kkkk (K)120  
222: kkkk (K)120  
223: kkkk (K)120  
224: kkkk (K)120  
225: kkkk (K)120

226: KKKKK (K)(20)  
 KKKKKKKKKKKKKKKKK KNRHH  
 227: KKKKK (K)(20)  
 KKKKKKKKKKKKKKKKK NRHHI

035807 ck: 7510 len: 129 i 035807 rattus norvegicus (rat). microvascu  
 (R,K)(20)  
 85: VLLAS (K)(20)  
 KKKKKKKKKKKKKKKKK KIKWE  
 86: LASK (K)(20)  
 KKKKKKKKKKKKKKKKK IKWEG

064075 ck: 8048 len: 215 i 064075 rattus sp. nucleoporin p62 homolog F  
 (R,K)(20)  
 35: CEFLE (K)(20)  
 KKKKKKKKKKKKKKKKK KKTGD  
 (K)(20)  
 36: EFLEK KKKKKKKKKKKKKKKKK KTGDN  
 (K)(20)  
 37: FLEKK KKKKKKKKKKKKKKKKK TGDNA

Databases searched:  
 SWISS-PROT, Release 40.1, Released on 11Feb2002, Formatted on 19Feb2002  
 SPTREMBL, Release 19.0, Released on 14Dec2001, Formatted on 19Feb2002

Total finds: 286  
 Total length: 211,714,479  
 Total sequences: 667,446  
 CPU time: 18:32.89

!!SEQUENCE LIST 1.0  
! FINDPATTERNS on swp:\* allowing 0 mismatches

1 (R,K){20}

July 1, 2002 15:28 ..

SP_FUN:Q12444	ck: 1384	len: 126	finds: 3	! Q12444	saccharomyces cerevisiae
SP_FUN:Q9P529	ck: 291	len: 128	finds: 13	! Q9P529	neurospora crassa . hypc
SP_HUM:Q9HC48	ck: 7602	len: 667	finds: 1	! Q9HC48	homo sapiens (human) . c
SP_HUM:Q9H6Q7	ck: 3351	len: 720	finds: 10	! Q9H6Q7	homo sapiens (human) . c
SP_HUM:Q9H5V6	ck: 379	len: 168	finds: 10	! Q9H5V6	homo sapiens (human) . c
SP_HUM:Q9NT34	ck: 7330	len: 380	finds: 7	! Q9NT34	homo sapiens (human) . h
SP_OM:Q95LV6	ck: 7515	len: 531	finds: 11	! Q95LV6	macaca fascicularis (cd
SP_PL:Q9LXR2	ck: 4143	len: 517	finds: 4	! Q9LXR2	arabidopsis thaliana (m
SP_PL:Q9LG29	ck: 6094	len: 260	finds: 222	! Q9LG29	arabidopsis thaliana (m
SP_RO:Q35807	ck: 7510	len: 129	finds: 2	! Q35807	rattus norvegicus (rat)
SP_RO:Q64075	ck: 8048	len: 215	finds: 3	! Q64075	rattus sp. nucleoporin

\\End of list

Databases searched:

SWISS-PROT, Release 40.1, Released on 11Feb2002, Formatted on 19Feb2002  
SPTRMBL, Release 19.0, Released on 14Dec2001, Formatted on 19Feb2002

Total finds: 286  
Total length: 211,714,479  
Total sequences: 667,446  
CPU time: 21:25.76

```
11AA_SEQUENCE 1.0
ID 012444 PRELIMINARY; PRT: 126 AA.
AC 012444;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
ORF YOR309C.
GN YOR309C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pearson B.M., Hernandez Y., Kalogetopoulos A., Schweizer M.;
Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MIPS;
Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN-FY1679;
Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Pearson B.M., Hernandez Y., Wolf S.S., Kalogetopoulos A., Schweizer M.;
Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z75217; CAAG9629.1; -.
DR EMBL: X90565; CAA62164.1; -.
DR SCD: S0005836; YOR309C.
SQ SEQUENCE 126 AA; 16294 MM; 46EF1F4C664802C8 CRC64;
012444; Length: 126 July 1, 2002 15:52 Type: P Check: 1384 ..

1 MOWLIPORL LILNPLIMK RKKRKKRKR RERETMKIP RILKLRKR
51 RTRRRRRKR KRRRRKRKR RKRSPKRKR KRRNKDAFYI LIIDPSRL
101 LFGFRKFSII IQCLTFSPH ILFHNL

11AA_SEQUENCE 1.0
ID 09P529 PRELIMINARY; PRT: 128 AA.
AC 09P529;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHEICAL 15.2 KDA PROTEIN.
GN B24H17.160.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Farlmann B., Holland R.,
Nyekatura G., Mewes H.W., Mannhaupt G.;
Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL356815; CAB92638.2; -.
KM Hypothetical protein.
SQ SEQUENCE 128 AA; 15157 MM; 8C7C65C3DFB70765 CRC64;
09P529 Length: 128 July 1, 2002 15:52 Type: P Check: 291 ..

1 MAIISGLHH KNINRAPGH SVYKSSSYD FQOYDATOHY LPSGFKAI
51 DHLGKGDC LSHDKRKNQ KKKKKKKKK KKKKKKKKK KKKKKKKKK
101 KKEESRTYF QOHFQADGIC PRPEWHTR

11AA_SEQUENCE 1.0
ID 09HC48 PRELIMINARY; PRT: 667 AA.
```

```
AC 09HC48;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CTCL TUMOR ANTIGEN SE2-5 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-TESTIS.;
MEDLINE-21143360; PubMed-1149944;
RA Eichmuller S., Usener D., Dummer R., Stein A., Thiel D.,
Schadendorf D.;
Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RT "Serological detection of cutaneous T-cell lymphoma-associated
antigens."
RT Proc. Natl. Acad. Sci. U.S.A. 98:629-634(2001).
DR EMBL: AF177228; AAC33676.1; -.
DR HSSP: Q12923; 3PDZ.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 2.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
FT NON_TER 1
FT TER 667
SQ SEQUENCE 667 AA; 73499 MM; C653EC16802BAE02 CRC64;
09HC48 Length: 667 July 1, 2002 15:52 Type: P Check: 7602 ..

1 EHENLFREND CIVRINDGL RNRFFEOAH MFRQAMRTPI IMFHVVPAN
51 KQEQYLSQS EKNYYSRF SPDSQYIDNR SVNSAGLHYV QRAPLHNPP
101 EQIDSHSRIP HSAHPGCKP SAPASAPQV FSTYVSSGN TKKIKRLNI
151 QLKKGTEGLE FSITSRDYTI GGSAPIYKN ILPGCAIAD GLKAGDRLI
201 EYNGVDLVGK SOEEVSLLR STKMEGTSL LVFRQDAPH PRELNAEPSQ
251 MQIPRETKAE DEDIVLPDG TREFLTFEVP LMDSGAGAG VSKNRNKE
301 NHADIGIFVK SIINGAASK DGLRLVNDL IAVNGESLLG KTNODAMETL
351 RRSMTGEGNK RGMQLIYAR RISKCNELKS PCSPGPELP IETALDRER
401 RISHLSYGI EGLDESPSRN AALSRIKES GKYQLSPYV MPODTVILE
451 DDRLPVLPPH LSDOSSSSSH DDVGFYTADA GTWAKAATSD SADCSLSPDV
501 DPVLAFOREG FGRQIADETK LNTVDQKAG SPSRDVPSL GLKSSSLES
551 LQTAVAEYTL NGDLPFHRPR PRILIRGCN ESFRAIDKS YDKPAVDDDD
601 BGMETLEEDT EESSRSGRES VSTAADPSH SLERQNGNO EKGDKTRKK
651 KKKKKKKKK KKKKKKK

11AA_SEQUENCE 1.0
ID 09H607 PRELIMINARY; PRT: 720 AA.
AC 09H607;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CDNA: FLJ21979 FIS, CLONE HEP06065 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hiki J. T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
Okitani R., Ota T., Suzuki Y., Odayashi M., Nishi T., Shibahara T.,
```

RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
 RT "MEDO human cDNA sequencing project."  
 RL Submitted (JUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK025632; BAB15196.1; -  
 FT NON\_TER 720 720  
 SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;

Q9H607 Length: 720 July 1, 2002 15:52 Type: P Check: 3351 ..

1 MLTEQVEQYT KEMKNTCII EDLKNELQKN KGASTLSQOT HMKIQSTLIDI  
 51 LKEKTKAER TAELEADAR EKDELVEAL KRLKDESGV YGLEDAVEI  
 101 KNCKNQIKIR DREIELITKE INKLELKISD FLDENEALRE RVGLEPKTMI  
 151 DLTEFRSKH LKQOYRAEN QILLKEIESL EERERDLKK IRQAQERGK  
 201 RSATSGLTTE DLNLTENISQ GDRISERKID LLSLKNMSEA QSKNEPLSRE  
 251 LIEKERDLER SRTVIKFNQ KLELVEENK QLEEGMKEIL QALIKMOKDP  
 301 DVKGETSLI IPSLERLVNA IESKNAEGIF DASLHLKAQV DQLTGRNDEL  
 351 ROELRESKKE AINYSQOLAK ANLKIDHLEK ETSLLROSEG SNVYFKGIDL  
 401 PDGIAPSSAS IINSQNEYL ILLQLEENKE KKLKNLEDSL EDYNRKFAVI  
 451 RHQOSLLYKE YLSEKETWKT ESKTIKEKR KLEDQVQDA IKVKEYNNLL  
 501 NALQMSDEM KKLAEKSRK ITVLOYNEKS LIRQYTTLVE LERLQRENE  
 551 KQKNELLME AEVCEKIGCL QRFKEMAFK IALQKVVDN SVSLSELELA  
 601 NKQYNELTAK YRDILQKDN LVQRTSNLEH LECENISLKE QVESINKELF  
 651 ITRKELHTIE QAMDEETKLG NESSMDKAKK SITNSDIYSI SKKKKKKKKK  
 701 KKKKKKKKK KKKKKKKKK

11AA\_SEQUENCE 1.0  
 ID Q9H5V6 PRELIMINARY; PRT: 168 AA.  
 AC Q9H5V6:  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE CNA: FLJ22976 FIS, CLONE KAT11222 (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
 RA Suzuki Y., Ohtsushiki M., Nishi T., Shibahara T., Tanaka T.,  
 RA Nakamura Y., Isogai T., Sugano S.;  
 RT "MEDO human cDNA sequencing project."  
 RL Submitted (JUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK026629; BAB15513.1; -  
 FT NON\_TER 168 168  
 SQ SEQUENCE 168 AA; 19549 MW; A19BD195F8A1A90 CRC64;

Q9H5V6 Length: 168 July 1, 2002 15:52 Type: P Check: 379 ..

1 MGNGRGSLQ QGKGNVQVA ATPTAASAC QYRCIECNOE AKELYRDYNNH  
 51 GVLITICKS CQKPVKYE YDPVILINA ILCKAQAVRH ILFTQIINIH  
 101 GKLCIFCLIC EAYLRWQLO DSNQNTAPD LIRYREWEK KKKKKKKKK  
 151 KKKKKKKKK KKKKKKKKK

11AA\_SEQUENCE 1.0  
 ID Q9NT34 PRELIMINARY; PRT: 380 AA.

AC Q9NT34:  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE HYPOTHETICAL 42.7 KDA PROTEIN (FRAGMENT).  
 GN DKF2P4341120.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Ohtenaeider B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL137556; CAB70810.1; -  
 KW Hypothetical protein.  
 FT NON\_TER 380 380  
 SQ SEQUENCE 380 AA; 42689 MW; 67F50DD101346AFB CRC64;

Q9NT34 Length: 380 July 1, 2002 15:52 Type: P Check: 7330 ..

1 MGSTDSKLNK KRAVIQLTTK TOPVEATDDA FWDQFWADTA TSVDVFAVL  
 51 PAEIRAVRE ESPSNLALIC YKAVEKLYOG AESGCHSEKE KQIVLNC SRL  
 101 LTRVLPIYFE DPMWRGFWS TVPGACRGG EEDDEHARL AESLLAIAD  
 151 LLPCDFITVQ SHRSYVDSA EDVHSLDCE YIWEAGVGA HSPDPNIHD  
 201 MNRMLKLKL LTCFSEAMYL PPAPESGSTN PWVQFFCSTE NRHALPLFTS  
 251 LLNTVCAYDP VGYGIPYNNH LFSDYREPLY EEAQVLIYV LDHDSASAS  
 301 PTVDGTTTGT AMDADPRGP ENLFVNYLSR IHRDFQFI LKGIARLLSN  
 351 LLLQKKKKKK KKKKKKKKK

11AA\_SEQUENCE 1.0  
 ID Q95LV6 PRELIMINARY; PRT: 531 AA.  
 AC Q95LV6:  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE HYPOTHETICAL 61.4 KDA PROTEIN (FRAGMENT).  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
 OC Cercopitheidae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,  
 RA Terao K., Sugano S.;  
 RT "Isolation of novel full-length cDNA clones from macaque testis cDNA  
 RT libraries."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB071085; BAB64479.1; -  
 KW Hypothetical protein.  
 FT NON\_TER 531 531  
 SQ SEQUENCE 531 AA; 61389 MW; B55996BAF5CDD60C CRC64;

Q95LV6 Length: 531 July 1, 2002 15:52 Type: P Check: 7515 ..

1 MESESSNAN MNVQHEREDK NIQKLPESV PCYSOHLFS TYQKKDPDPC  
 51 KSRSEPKSPE GRSSWNLSHI VQTEQETHF RESVLEPISG YMMQSPHMQ  
 101 EGICVGVGLK TSFPTKCKSE ICSMPHDTPW DENPRKKDS SISKTAMP  
 151 KNLOTVLKPL DFSSLSMSSEY ESRSTYLEFT GKKSITSPKH VTLKTOLPI

201 SOLFNINCS TENHKKKH CFTYKMGRO WYTSIDEALR SATEYAKSP  
251 SKSMIDKLF NTAARGLSN RTHQNYGH TTEEKEVOE NVAASLGL  
301 DEFMVLDS KNQNTIRLS ERKTLNPK LTKKESKSI SOIRKINH  
351 TTKHKKKLS NLKTKLMM QGENVDTFP NTSFPDTS DIKROSPOT  
401 EIDMRISGLS HTOPTQJESL AEGJARCSK RRTSNLVKGT KLHRESGK  
451 KOEHLTGMDP FYAENFMNT HLKRDPLHG SEDVLLGEFP ISKSOFYKGN  
501 SKKKKKKKK KKKKKKKKK KKKKKKKKK K

11AA\_SEQUENCE 1.0  
ID 09LGR2 PRELIMINARY: PRT: 517 AA.

AC 09LGR2: 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
GN HYPOHETICAL 59.7 KDA PROTEIN.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxId=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,  
RA Rued S., Lemcke K., Mayer K.F.X., Quettier F., Salanoubat M.,  
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project.  
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL353032; CAB8307.1; -  
DR InterPro: IPR001810; F-box.  
DR Pfam: PF00646; F-box; 1.  
DR SMART: SM00256; FBOX; 1.  
DR PROSITE: PS50181; FBOX; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 517 AA; 59689 MW; EC6D957D01F86E70 CRC64;

09LGR2 Length: 517 July 1, 2002 15:52 Type: P Check: 4143

1 MDLFSLPNE LLYHLSLS TKEALTSVL SKRWNLFAF VPYLEFDDSV  
51 FLHPEERRE KEGILQSFMD FVDRVLDLHG DSLIKTFSLK CTGVDSDHV  
101 DRWICNVLAR GVSDDLFLID FRDLYSLPHE VGVSRLLVYL RVGSSESLYV  
151 WOKFLCLPML KTVLVDSGL CIGOFOLLL ACPALELDM TNRKWKDSNV  
201 TVSSSILKEL TIDLHGCCSV VNLKSLFDA PSIVYFYVCD SLAEDYPOVN  
251 LKNLVEAQLN LLLTQAQIEQ VRALNNEMLV ADDVFPGLON AMKLITGLRN  
301 VOQLYLSPT LEVLSRCCEG MPVFNMLKVL STWSDNNRCW QAMPVLLRNC  
351 PHLETLLEG LLAHYATDKG DVCDCISRDY KDHSLSLCPV KKLQYEFRG  
401 TIRELEMKH FLKIFPCLKE MDIYAHENSH TLFKOPTIFE RVGKKKKKKK  
451 KKKKKKKKK KKKKKKTRLN FKPVNTEOP LKRLADKICE IPQCLEFLDV  
501 DSSLGELALL AMDSRPS

11AA\_SEQUENCE 1.0  
ID 09LGR2 PRELIMINARY: PRT: 260 AA.  
AC 09LGR2: 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE: F1D9.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxId=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-COLUMBIA;  
RA Nakamura Y.;  
RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AP002460; BAA97098.1; -  
DR InterPro: IPR001386; Linker-histone.  
DR PRINTS: PR00624; HISTONEH5.  
SQ SEQUENCE 260 AA; 33307 MW; 43E2394CB8131143 CRC64;

09LGR2 Length: 260 July 1, 2002 15:52 Type: P Check: 6094

1 MDRCIKRRKK KKKKKKKKK KKKKKKKKK KKKKKKKKK  
51 KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK  
101 KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK  
151 KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK  
201 KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK  
251 INKMGFVIF

11AA\_SEQUENCE 1.0  
ID 035807 PRELIMINARY: PRT: 129 AA.  
AC 035807:  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE MICROVASCULAR ENDOTHELIAL DIFFERENTIATION PROTEIN 2.  
GN MDG2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxId=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EPIDIDYMIS;  
RX MEDLINE=98172708; PubMed=9511718;  
RA Proels F., Loser B., Marx M.;  
RT "Differential expression of osteopontin, PC4, and CEC5, a novel mRNA  
RT species, during in vitro angiogenesis.";  
RL Exp. Cell Res. 239:1-10(1998).  
DR EMBL: Y08769; CAA70022.1; -  
DR InterPro: IPR000719; Euk.pkinase.  
DR Pfam: PF00069; pkinase; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW ATP-binding; Transferase.  
SQ SEQUENCE 129 AA; 15080 MW; 38102272BBE2EDB4 CRC64;

035807 Length: 129 July 1, 2002 15:52 Type: P Check: 7510

1 MKKHPIVEL LETYSSDGL YNVEFMDGA DLCEFIYKRA DAGVYSEAV  
51 ASHYMKOILE ALRYCHDNNI IHRDVPRHCY LLSKKKKKK KKKKKKKKK  
101 KKKKKKMEG RDAFWAIPV KSSGCVIIO

11AA\_SEQUENCE 1.0  
ID 064075 PRELIMINARY: PRT: 215 AA.  
AC 064075:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)



DE NUCLEOPORIN P62 HOMOLOG PROTEIN (FRAGMENT).  
OS Rattus sp.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10118;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95151924; Pubmed=7849178;  
RA Wang Z.O., Akmal K.M., Kim K.H.;  
RT "An unusual nucleoporin-related messenger ribonucleic acid is present  
in the germ cells of rat testis.";  
RL Biol. Reprod. 51:1022-1030(1994).  
DR EMBL: S75997; AAB33384.1; -  
KW Porin.  
FT NON\_TER 1 1  
SQ SEQUENCE 215 AA: 24593 MW: 098251C97A8FBD88 CRC64:  
Q64075 Length: 215 July 1, 2002 15:52 Type: P Check: 8048 ..  
1 SGRATSSCD EDCLSSSLPF SLGPKVQDC EFLEKKKKK KKKKKKKK  
51 KKKKKKTGDN AKSVRQYSL KTKLEHAE QAKVELDFTL SQKLEDDL  
101 SPLESVKEQ SGTIYLOHAD EEREKTYKLA ENIDAQLKRM AQDLKDIIH  
151 LNMAGPADT SDPLQIQICKI LNAHMDSLQW VDSSALLQK RVEEASRYCE  
201 SRKKEQERSL RIAFD

! FINDPATTERNS on geneseqp.\* allowing 0 mismatches

1 (R,K){20} July 1, 2002 06:23 ..

1 AAP20159 ck: 5750 len: 20 ! Aap20159 Sequence of lysine polymer. 8/1992

1: (R,K){20}  
(K){20}  
XXXXXXXXXXXXXXXXXXXX

1 AAP61030 ck: 9157 len: 898 ! Aap61030 Entire coded sequence from plasmid

873: KNITW (R,K){20}  
(K){20} XXXXXXXXXXXXXXXXXXXX

874: NITWK (K){20} XXXXXXXXXXXXXXXXXXXX

875: ITWKK (K){20} XXXXXXXXXXXXXXXXXXXX

876: TWKKK (K){20} XXXXXXXXXXXXXXXXXXXX

877: WKKKK (K){20} XXXXXXXXXXXXXXXXXXXX

878: KKKKK (K){20} XXXXXXXXXXXXXXXXXXXX

879: KKKKK (K){20} XXXXXXXXXXXXXXXXXXXX

1 AAP61056 ck: 2017 len: 899 ! Aap61056 Translation of plasmid PAU157 endc

873: KNITW (R,K){20} XXXXXXXXXXXXXXXXXXXX

874: NITWK (K){20} XXXXXXXXXXXXXXXXXXXX

875: ITWKK (K){20} XXXXXXXXXXXXXXXXXXXX

876: TWKKK (K){20} XXXXXXXXXXXXXXXXXXXX

877: WKKKK (K){20} XXXXXXXXXXXXXXXXXXXX

878: KKKKK (K){20} XXXXXXXXXXXXXXXXXXXX

879: KKKKK (K){20} XXXXXXXXXXXXXXXXXXXX

880: KKKKK (K){20} XXXXXXXXXXXXXXXXXXXX

1 AAP61082 ck: 7915 len: 898 ! Aap61082 Complete translation of plasmid PA

873: KNITW (R,K){20} XXXXXXXXXXXXXXXXXXXX

874: NITWK (K){20} XXXXXXXXXXXXXXXXXXXX

875: ITWKK (K){20} XXXXXXXXXXXXXXXXXXXX

876: TWKKK (K){20} XXXXXXXXXXXXXXXXXXXX

877: WKKKK (K){20} XXXXXXXXXXXXXXXXXXXX

878: KKKKK (K){20} XXXXXXXXXXXXXXXXXXXX

879: KKKKK (K){20} XXXXXXXXXXXXXXXXXXXX

1 AAR29580 ck: 4341 len: 657 ! Aar29580 FMR-1 gene product. 4/1993

19: RRRRP (R,K){20} RRRRRRRRRRRRRRRRRR

20: RRRPR (R){20} RRRRRRRRRRRRRRRRRR

21: RRP RR (R){20} RRRRRRRRRRRRRRRRRR

22: RPRRR (R){20} RRRRRRRRRRRRRRRRRR

23: PRRRR (R){20} RRRRRRRRRRRRRRRRRR

1 AAW03642 ck: 9623 len: 116 ! Aaw03642 Human cannabinoid GPR N-termina

34: QYEDI (R,K){20} XXXXXXXXXXXXXXXXXXXX

35: YEDIK (K){20} XXXXXXXXXXXXXXXXXXXX

1 AAW38839 ck: 801 len: 28 ! Aaw38839 Delivery peptide used in peptid

1: (R,K){20} XXXXXXXXXXXXXXXXXXXX

2: K (K){20} XXXXXXXXXXXXXXXXXXXX

3: KK (K){20} XXXXXXXXXXXXXXXXXXXX

4: KKK (K){20} XXXXXXXXXXXXXXXXXXXX

5: KKKK (K){20} XXXXXXXXXXXXXXXXXXXX

6: KKKKK (K){20} XXXXXXXXXXXXXXXXXXXX

7: KKKKK (K){20} XXXXXXXXXXXXXXXXXXXX

1 AAW38840 ck: 2989 len: 29 ! Aaw38840 Delivery peptide used in peptid

1

```
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKK XK
```

AAW38841 ck: 5252 len: 30 i Aaw38841 Delivery peptide used in peptide m

1

```
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKK KKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKK XK
```

AAW38842 ck: 7590 len: 31 i Aaw38842 Delivery peptide used in peptide m

1

```
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
```

1

```
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKK KKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKK XK
```

AAW38877 ck: 1129 len: 23 i Aaw38877 Delivery peptide used in pepti

1

```
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK VTK
```

AAW38843 ck: 3 len: 32 i Aaw38843 Delivery peptide used in pepti

1

```
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKK KKK
(K){20}
11: KKKKK KKKKKKKKKKKKKKKKK XK
```

AAW38878 ck: 2949 len: 24 i Aaw38878 Delivery peptide used in pepti

1

```
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK KVTK
(K){20}
2: K KKKKKKKKKKKKKKKKK VTK
```

AAW38844 ck: 2491 len: 33 i Aaw38844 Delivery peptide used in pepti

1

(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKK KKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKK XK

1

AAW38879 ck: 4844 len: 25 ! Aaw38879 Delivery peptide used in peptide n  
(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKK KKVTK  
(K){20}  
2: K KKKKKKKKKKKKKKKKK KVTK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKK VTK

1

AAW38845 ck: 5054 len: 34 ! Aaw38845 Delivery peptide used in peptide n  
(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK

1

(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKK KKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKK XK

1

AAW38846 ck: 7692 len: 35 ! Aaw38846 Delivery peptide used in peptid  
(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKK KKKKK

```
12: KKKKK (K){20} KKKKKKKKKKKKKKKKK KKKK
13: KKKKK (K){20} KKKKKKKKKKKKKKKKK KKK
14: KKKKK (K){20} KKKKKKKKKKKKKKKKK XK

AAW38833 ck: 9248 len: 22 i Aaw38833 Delivery peptide used in peptide m
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKK KKK

AAW38834 ck: 986 len: 23 i Aaw38834 Delivery peptide used in peptide m
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKK KKK
2: K (K){20} KKKKKKKKKKKKKKKKKKKKK XK

AAW38835 ck: 2799 len: 24 i Aaw38835 Delivery peptide used in peptide m
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKK KKK
2: K (K){20} KKKKKKKKKKKKKKKKKKKKK KKK
3: K (K){20} KKKKKKKKKKKKKKKKKKKKK XK

AAW38836 ck: 4687 len: 25 i Aaw38836 Delivery peptide used in peptide m
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKK KKK
2: K (K){20} KKKKKKKKKKKKKKKKKKKKK KKK
3: K (K){20} KKKKKKKKKKKKKKKKKKKKK KKK
4: K (K){20} KKKKKKKKKKKKKKKKKKKKK XK

AAW38837 ck: 6650 len: 26 i Aaw38837 Delivery peptide used in peptide m
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKK KKK
2: K (K){20} KKKKKKKKKKKKKKKKKKKKK KKK
3: K (K){20} KKKKKKKKKKKKKKKKKKKKK KKK
4: K (K){20} KKKKKKKKKKKKKKKKKKKKK KKK

5: KKKKK KKKKKKKKKKKKKKKKKKKKKKK XK
AAW38838 ck: 8688 len: 27 i Aaw38838 Delivery peptide used in peptide m
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKK KKK
2: K (K){20} KKKKKKKKKKKKKKKKKKKKK KKK
3: K (K){20} KKKKKKKKKKKKKKKKKKKKK KKK
4: K (K){20} KKKKKKKKKKKKKKKKKKKKK KKK
5: K (K){20} KKKKKKKKKKKKKKKKKKKKK KKK
6: K (K){20} KKKKKKKKKKKKKKKKKKKKK XK

AAW38796 ck: 9227 len: 22 i Aaw38796 Delivery peptide used in peptide m
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKK WK

AAW38797 ck: 964 len: 23 i Aaw38797 Delivery peptide used in peptide m
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKK KWK
2: K (K){20} KKKKKKKKKKKKKKKKKKKKK WK

AAW38798 ck: 2776 len: 24 i Aaw38798 Delivery peptide used in peptide m
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKK KWK
2: K (K){20} KKKKKKKKKKKKKKKKKKKKK KWK
3: K (K){20} KKKKKKKKKKKKKKKKKKKKK WK

AAW38799 ck: 4663 len: 25 i Aaw38799 Delivery peptide used in peptide m
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKK KWK
2: K (K){20} KKKKKKKKKKKKKKKKKKKKK KWK
3: K (K){20} KKKKKKKKKKKKKKKKKKKKK KWK
4: K (K){20} KKKKKKKKKKKKKKKKKKKKK WK

AAW38800 ck: 6625 len: 26 i Aaw38800 Delivery peptide used in peptide m
```

1  
1: (R,K){20}  
KKKKKKKKKKKKKKKKKKKK  
2: (K){20}  
K KKKKKKKKKKKKKKKKKKK  
3: (K){20}  
KK KKKKKKKKKKKKKKKKKKK  
4: (K){20}  
KKK KKKKKKKKKKKKKKKKKKK  
5: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK WK

AAW38801 ck: 8662 len: 27 i Aaw38801 Delivery peptide used in peptide m  
1: (R,K){20}  
(K){20}  
KKKKKKKKKKKKKKKKKKKK  
2: (K){20}  
K KKKKKKKKKKKKKKKKKKK  
3: (K){20}  
KK KKKKKKKKKKKKKKKKKKK  
4: (K){20}  
KKK KKKKKKKKKKKKKKKKKKK  
5: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
6: (K){20}  
KKKKK KKKKKKKKKKKKKKKKKKK WK

AAW38802 ck: 774 len: 28 i Aaw38802 Delivery peptide used in peptide m  
1: (R,K){20}  
(K){20}  
KKKKKKKKKKKKKKKKKKKK  
2: (K){20}  
K KKKKKKKKKKKKKKKKKKK  
3: (K){20}  
KK KKKKKKKKKKKKKKKKKKK  
4: (K){20}  
KKK KKKKKKKKKKKKKKKKKKK  
5: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
6: (K){20}  
KKKKK KKKKKKKKKKKKKKKKKKK  
7: (K){20}  
KKKKK KKKKKKKKKKKKKKKKKKK WK

AAW38803 ck: 2961 len: 29 i Aaw38803 Delivery peptide used in peptide m  
1: (R,K){20}  
(K){20}  
KKKKKKKKKKKKKKKKKKKK  
2: (K){20}  
K KKKKKKKKKKKKKKKKKKK

3: (K){20}  
KK KKKKKKKKKKKKKKKKKKK  
4: (K){20}  
KKK KKKKKKKKKKKKKKKKKKK  
5: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
6: (K){20}  
KKKKK KKKKKKKKKKKKKKKKKKK  
7: (K){20}  
KKKKK KKKKKKKKKKKKKKKKKKK  
8: (K){20}  
KKKKK KKKKKKKKKKKKKKKKKKK WK

AAW38804 ck: 5223 len: 30 i Aaw38804 Delivery peptide used in peptide m  
1: (R,K){20}  
(K){20}  
KKKKKKKKKKKKKKKKKKKK  
2: (K){20}  
K KKKKKKKKKKKKKKKKKKK  
3: (K){20}  
KK KKKKKKKKKKKKKKKKKKK  
4: (K){20}  
KKK KKKKKKKKKKKKKKKKKKK  
5: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
6: (K){20}  
KKKKK KKKKKKKKKKKKKKKKKKK  
7: (K){20}  
KKKKK KKKKKKKKKKKKKKKKKKK  
8: (K){20}  
KKKKK KKKKKKKKKKKKKKKKKKK  
9: (K){20}  
KKKKK KKKKKKKKKKKKKKKKKKK WK

AAW38805 ck: 7560 len: 31 i Aaw38805 Delivery peptide used in peptide m  
1: (R,K){20}  
(K){20}  
KKKKKKKKKKKKKKKKKKKK  
2: (K){20}  
K KKKKKKKKKKKKKKKKKKK  
3: (K){20}  
KK KKKKKKKKKKKKKKKKKKK  
4: (K){20}  
KKK KKKKKKKKKKKKKKKKKKK  
5: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
6: (K){20}  
KKKKK KKKKKKKKKKKKKKKKKKK  
7: (K){20}  
KKKKK KKKKKKKKKKKKKKKKKKK

8: KKKK (K){20} KKKK  
9: KKKK (K){20} KKKK  
10: KKKK (K){20} KKKK

AAW38806 ck: 9972 len: 32 ! Aaw38806 Delivery peptide used in peptide m

1: (R,K){20} KKKK  
2: K (K){20} KKKK  
3: K (K){20} KKKK  
4: K (K){20} KKKK  
5: K (K){20} KKKK  
6: K (K){20} KKKK  
7: K (K){20} KKKK  
8: K (K){20} KKKK  
9: K (K){20} KKKK  
10: K (K){20} KKKK  
11: K (K){20} KKKK

AAW38807 ck: 2459 len: 33 ! Aaw38807 Delivery peptide used in peptide m

1: (R,K){20} KKKK  
2: K (K){20} KKKK  
3: K (K){20} KKKK  
4: K (K){20} KKKK  
5: K (K){20} KKKK  
6: K (K){20} KKKK  
7: K (K){20} KKKK

8: KKKK (K){20} KKKK  
9: KKKK (K){20} KKKK  
10: KKKK (K){20} KKKK  
11: KKKK (K){20} KKKK  
12: KKKK (K){20} KKKK

AAW38808 ck: 5021 len: 34 ! Aaw38808 Delivery peptide used in peptide m

1: (R,K){20} KKKK  
2: K (K){20} KKKK  
3: K (K){20} KKKK  
4: K (K){20} KKKK  
5: K (K){20} KKKK  
6: K (K){20} KKKK  
7: K (K){20} KKKK  
8: K (K){20} KKKK  
9: K (K){20} KKKK  
10: K (K){20} KKKK  
11: K (K){20} KKKK  
12: K (K){20} KKKK  
13: K (K){20} KKKK

AAW38881 ck: 8859 len: 27 ! Aaw38881 Delivery peptide used in peptide m

1: (R,K){20} KKKK  
2: K (K){20} KKKK  
3: K (K){20} KKKK  
4: K (K){20} KKKK

5: KKKK (K){20} VTK

AAW38847 ck: 405 len: 36 ! Aaw38847 Delivery peptide used in peptide m

(R,K){20}

(K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

AAW38882 ck: 979 len: 28 ! Aaw38882 Delivery peptide used in peptide m

(R,K){20}

(K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

AAW38848 ck: 3193 len: 37 ! Aaw38848 Delivery peptide used in peptid

(R,K){20}

(K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

AAW38883 ck: 3174 len: 29 ! Aaw38883 Delivery peptide used in peptid

(R,K){20}

(K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}



6: KKKKK KKKKKKKKKKKKKKKKKKK KVTK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKK VTK

AAW38849 ck: 6056 len: 38 1 Aaw38849 Delivery peptide used in peptide m

1

(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
2: K KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK

(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK

(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KK

AAW38884 ck: 5444 len: 30 1 Aaw38884 Delivery peptide used in peptide m

1

(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
2: K KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
3: KK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKVT

(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKVTK

(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKKKK KVTK

(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKKKK VTK

AAW38850 ck: 8994 len: 39 1 Aaw38850 Delivery peptide used in peptl.

1

(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
2: K KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
3: KK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKK

(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKK

(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KK

```
1
1 AAW38885 ck: 7789 len: 31 i Aaw38885 Delivery peptide used in peptide n
(R,K)[20]
(K)[20]
1: KKKKKKKKKKKKKKKKKKKKK KKKKK
2: K KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
3: KK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
7: KKKKK KKKKKKKKKKKKKKKKKKK KKVTK
(K)[20]
8: KKKKK KKKKKKKKKKKKKKKKKKK KVTK
(K)[20]
9: KKKKK KKKKKKKKKKKKKKKKKKK VTK
AAW38851 ck: 2007 len: 40 i Aaw38851 Delivery peptide used in peptide n
(R,K)[20]
(K)[20]
1: KKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
2: K KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
3: KK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
```

```
1
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
AAW38886 ck: 209 len: 32 i Aaw38886 Delivery peptide used in peptide n
(R,K)[20]
(K)[20]
1: KKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
2: K KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
3: KK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
7: KKKKK KKKKKKKKKKKKKKKKKKK KKVTK
(K)[20]
8: KKKKK KKKKKKKKKKKKKKKKKKK KKVTK
(K)[20]
9: KKKKK KKKKKKKKKKKKKKKKKKK KVTK
(K)[20]
10: KKKKK KKKKKKKKKKKKKKKKKKK VTK
AAW38852 ck: 5095 len: 41 i Aaw38852 Delivery peptide used in peptide n
(R,K)[20]
(K)[20]
1: KKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
2: K KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
3: KK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
```

```

6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
19: KKKKK KKKKKKKKKKKKKKKKKKK KKK
   (K) (20)
20: KKKKK KKKKKKKKKKKKKKKKKKK KKK
   (K) (20)

AAW38887 ck: 2704 len: 33 i Aaw38887 Delivery peptide used in peptide m
   (R,K) (20)
   (K) (20)
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
5: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)

```

```

1
AAW38853 ck: 8258 len: 42 i Aaw38853 Delivery peptide used in peptid
   (R,K) (20)
   (K) (20)
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
5: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
16: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
17: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
18: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
19: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) (20)
20: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKK
   (K) (20)

```

21: KKKKK KKKKKKKKKKKKKKKKKKKKK XK

AAW38888 ck: 5274 len: 34 ! Aaw38888 Delivery peptide used in peptide m

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

AAW38854 ck: 1496 len: 43 ! Aaw38854 Delivery peptide used in peptide m

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

10: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

11: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

12: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

13: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

14: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

15: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

16: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

17: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

18: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

19: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

20: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

21: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

AAW38889 ck: 7919 len: 35 ! Aaw38889 Delivery peptide used in peptid

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

```
(K){20}
11: KKKK KKKKKKKKKKKKKKKKK KKVTK
    (K){20}
12: KKKK KKKKKKKKKKKKKKKKK KVTK
    (K){20}
13: KKKK KKKKKKKKKKKKKKKKK VTK

AAW3890 ck: 639 len: 36 i Aaw3890 Delivery peptide used in peptide m
1: (R,K){20}
   (K){20}
   KKKKKKKKKKKKKKKKKKKKK KKKK
2: (K){20}
   K KKKKKKKKKKKKKKKKKKK KKKK
3: (K){20}
   KK KKKKKKKKKKKKKKKKKKK KKKK
4: (K){20}
   KKK KKKKKKKKKKKKKKKKKKK KKKK
5: (K){20}
   KKK KKKKKKKKKKKKKKKKKKK KKKK
6: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
7: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
8: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
9: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
10: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
11: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKVTK
12: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKVTK
13: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KVTK
14: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK VTK

AAW3891 ck: 3434 len: 37 i Aaw3891 Delivery peptide used in peptide m
1: (R,K){20}
   (K){20}
   KKKKKKKKKKKKKKKKKKKKK KKKK
2: (K){20}
   K KKKKKKKKKKKKKKKKKKK KKKK
3: (K){20}
   KK KKKKKKKKKKKKKKKKKKK KKKK
4: (K){20}
   KKK KKKKKKKKKKKKKKKKKKK KKKK
5: (K){20}
   KKK KKKKKKKKKKKKKKKKKKK KKKK
```

```
(K){20}
6: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
7: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
8: KKKK KKKKKKKKKKKKKKKKKKK KKKK
9: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
10: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
11: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
12: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKVTK
13: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KVTK
14: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KVTK
15: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK VTK

AAW3892 ck: 6304 len: 38 i Aaw3892 Delivery peptide used in pepti
1: (R,K){20}
   (K){20}
   KKKKKKKKKKKKKKKKKKKKK KKKK
2: (K){20}
   K KKKKKKKKKKKKKKKKKKK KKKK
3: (K){20}
   KK KKKKKKKKKKKKKKKKKKK KKKK
4: (K){20}
   KKK KKKKKKKKKKKKKKKKKKK KKKK
5: (K){20}
   KKK KKKKKKKKKKKKKKKKKKK KKKK
6: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
7: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
8: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
9: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
10: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
11: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
12: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKKK
13: (K){20}
   KKKK KKKKKKKKKKKKKKKKKKK KKVTK
```

(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKK KKVTK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKK KVTK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKK VTK

AAW3893 ck: 9249 len: 39 i Aaw3893 Delivery peptide used in peptide m

1

(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
2: K KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
3: KK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKVT

(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKVTK

(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KVTK

(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK VTK

AAW3894 ck: 2269 len: 40 i Aaw3894 Delivery peptide used in peptide m

1

(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
3: KK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKVT

(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KKVTK

(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KVTK

(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK VTK

AAW3895 ck: 5364 len: 41 i Aaw3895 Delivery peptide used in peptid

1

(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
2: K KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
3: KK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
AAW38896 ck: 8534 len: 42 i Aaw38896 Delivery peptide used in peptide p  
(R,K)(20)  
1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
5: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
AAW38897 ck: 1779 len: 43 i Aaw38897 Delivery peptide used in peptide  
(R,K)(20)  
1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
5: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKV  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKVT  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKVTK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KVTK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKKKKKK VTK  
(K){20}

AAW38898 ck: 5099 len: 44 ! Aaw38898 Delivery peptide used in peptide m

1  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKV  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKVT  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKVTK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KVTK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKKKKKK VTK  
(K){20}

AAW38809 ck: 7658 len: 35 ! Aaw38809 Delivery peptide used in peptid

1  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKV  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKV  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKV  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKK WK  
(K){20}

AAW38810 ck: 370 len: 36 ! Aaw38810 Delivery peptide used in peptid

1  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}



2: K (K)(20) KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 3: KK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 4: KKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)

AAW38811 ck: 3157 len: 37 ! Aaw38811 Delivery peptide used in peptide π

1: (R,K)(20) KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 2: K KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 3: KK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 4: KKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)

10: KKKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)

AAW38812 ck: 6019 len: 38 ! Aaw38812 Delivery peptide used in peptide π

1: (R,K)(20) KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 2: K KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 3: KK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 4: KKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)  
 16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K)(20)

17: KKKK (K){20} KKKKKKKKKKKKKKKKK KKK

AAW38813 ck: 8956 len: 39 ! Aaw38813 Delivery peptide used in peptide m

1 (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK

2: K KKKKKKKKKKKKKKKKKKK KKKKK

3: KKKKKKKKKKKKKKKKKKKKK KKKKK

4: KKKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKKKKKKKKKKKKKKKKKK KKKKK

9: KKKKKKKKKKKKKKKKKKKKK KKKKK

10: KKKKKKKKKKKKKKKKKKKKK KKKKK

11: KKKKKKKKKKKKKKKKKKKKK KKKKK

12: KKKKKKKKKKKKKKKKKKKKK KKKKK

13: KKKKKKKKKKKKKKKKKKKKK KKKKK

14: KKKKKKKKKKKKKKKKKKKKK KKKKK

15: KKKKKKKKKKKKKKKKKKKKK KKKKK

16: KKKKKKKKKKKKKKKKKKKKK KKKKK

17: KKKKKKKKKKKKKKKKKKKKK KKKKK

18: KKKKKKKKKKKKKKKKKKKKK KKKKK

AAW38814 ck: 1968 len: 40 ! Aaw38814 Delivery peptide used in peptide m

1 (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK

2: K KKKKKKKKKKKKKKKKKKK KKKKK

3: KKKKKKKKKKKKKKKKKKKKK KKKKK

4: KKKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKKKKKKKKKKKKKKKKKK KKKKK

1

4: KKKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKKKKKKKKKKKKKKKKKK KKKKK

9: KKKKKKKKKKKKKKKKKKKKK KKKKK

10: KKKKKKKKKKKKKKKKKKKKK KKKKK

11: KKKKKKKKKKKKKKKKKKKKK KKKKK

12: KKKKKKKKKKKKKKKKKKKKK KKKKK

13: KKKKKKKKKKKKKKKKKKKKK KKKKK

14: KKKKKKKKKKKKKKKKKKKKK KKKKK

15: KKKKKKKKKKKKKKKKKKKKK KKKKK

16: KKKKKKKKKKKKKKKKKKKKK KKKKK

17: KKKKKKKKKKKKKKKKKKKKK KKKKK

18: KKKKKKKKKKKKKKKKKKKKK KKKKK

19: KKKKKKKKKKKKKKKKKKKKK KKKKK

AAW38815 ck: 5055 len: 41 ! Aaw38815 Delivery peptide used in peptid

1 (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK

2: K KKKKKKKKKKKKKKKKKKK KKKKK

3: KKKKKKKKKKKKKKKKKKKKK KKKKK

4: KKKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKKKKKKKKKKKKKKKKKK KKKKK

```

      8: KKKK (K){20} KKKK
      9: KKKK (K){20} KKKK
     10: KKKK (K){20} KKKK
     11: KKKK (K){20} KKKK
     12: KKKK (K){20} KKKK
     13: KKKK (K){20} KKKK
     14: KKKK (K){20} KKKK
     15: KKKK (K){20} KKKK
     16: KKKK (K){20} KKKK
     17: KKKK (K){20} KKKK
     18: KKKK (K){20} KKKK
     19: KKKK (K){20} KKKK
     20: KKKK (K){20} KKKK WK

AAW3816 ck: 8217 len: 42 ! Aaw3816 Delivery peptide used in peptide m
      1: (R,K){20} KKKK
      2: K (K){20} KKKK
      3: K (K){20} KKKK
      4: K (K){20} KKKK
      5: K (K){20} KKKK
      6: K (K){20} KKKK
      7: K (K){20} KKKK
      8: K (K){20} KKKK
      9: K (K){20} KKKK
     10: K (K){20} KKKK

```

```

      11: KKKK (K){20} KKKK
      12: KKKK (K){20} KKKK
      13: KKKK (K){20} KKKK
      14: KKKK (K){20} KKKK
      15: KKKK (K){20} KKKK
      16: KKKK (K){20} KKKK
      17: KKKK (K){20} KKKK
      18: KKKK (K){20} KKKK
      19: KKKK (K){20} KKKK
      20: KKKK (K){20} KKKK WK
      21: KKKK (K){20} KKKK WK

AAW3817 ck: 1454 len: 43 ! Aaw3817 Delivery peptide used in peptid
      1: (R,K){20} KKKK
      2: K (K){20} KKKK
      3: K (K){20} KKKK
      4: K (K){20} KKKK
      5: K (K){20} KKKK
      6: K (K){20} KKKK
      7: K (K){20} KKKK
      8: K (K){20} KKKK
      9: K (K){20} KKKK
     10: K (K){20} KKKK
     11: K (K){20} KKKK
     12: K (K){20} KKKK

```

```

14: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
15: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
16: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
17: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
18: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
19: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
20: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
21: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
22: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
23: KKKK KKKKKKKKKKKKKKKKKKK KKK
      (K){20}
24: KKKK KKKKKKKKKKKKKKKKKKK WK
      (K){20}

AAW21590 ck: 4875 len: 30 i Aaw21590 Antibiotic potentiating peptide
1
      (R,K){20}
1: KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
2: K KKKKKKKKKKKKKKKKK KKKK
      (K){20}
3: KK KKKKKKKKKKKKKKKKK KKKK
      (K){20}
4: KK KKKKKKKKKKKKKKKKK KKKK
      (K){20}
5: KKK KKKKKKKKKKKKKKKKK KKKK
      (K){20}
6: KKKK KKKKKKKKKKKKKKKKK KKKK
      (K){20}
7: KKKK KKKKKKKKKKKKKKKKK KKK
      (K){20}
8: KKKK KKKKKKKKKKKKKKKKK KKK
      (K){20}
9: KKKK KKKKKKKKKKKKKKKKK KK
      (K){20}
10: KKKK KKKKKKKKKKKKKKKKK K
      (K){20}
11: KKKK KKKKKKKKKKKKKKKKK
      (K){20}

AAW21591 ck: 5075 len: 434 i Aaw21591 Antibiotic potentiating peptide
1
      (R,K){20}

```

1: (K){20}  
KKKKKKKKKKKKKKKKKKKK  
2: (K){20}  
K KKKKKKKKKKKKKKKKKKK  
3: (K){20}  
KK KKKKKKKKKKKKKKKKKKK  
4: (K){20}  
KK KKKKKKKKKKKKKKKKKKK  
5: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
6: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
7: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
8: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
9: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
10: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
11: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
12: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
13: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
14: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
15: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
16: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
17: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
18: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
19: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
20: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
21: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
22: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
23: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
24: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK

25: KKKKK KKKKKKKKKKKKKKKKKKK  
26: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
27: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
28: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
29: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
30: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
31: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
32: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
33: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
34: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
35: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
36: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
37: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
38: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
39: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
40: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
41: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
42: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
43: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
44: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
45: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
46: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
47: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
48: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK  
49: (K){20}  
KKKK KKKKKKKKKKKKKKKKKKK

50: kkkk (K){20} kkkk  
51: kkkk (K){20} kkkk  
52: kkkk (K){20} kkkk  
53: kkkk (K){20} kkkk  
54: kkkk (K){20} kkkk  
55: kkkk (K){20} kkkk  
56: kkkk (K){20} kkkk  
57: kkkk (K){20} kkkk  
58: kkkk (K){20} kkkk  
59: kkkk (K){20} kkkk  
60: kkkk (K){20} kkkk  
61: kkkk (K){20} kkkk  
62: kkkk (K){20} kkkk  
63: kkkk (K){20} kkkk  
64: kkkk (K){20} kkkk  
65: kkkk (K){20} kkkk  
66: kkkk (K){20} kkkk  
67: kkkk (K){20} kkkk  
68: kkkk (K){20} kkkk  
69: kkkk (K){20} kkkk  
70: kkkk (K){20} kkkk  
71: kkkk (K){20} kkkk  
72: kkkk (K){20} kkkk  
73: kkkk (K){20} kkkk



74: kkkk (K){20} kkkk  
75: kkkk (K){20} kkkk  
76: kkkk (K){20} kkkk  
77: kkkk (K){20} kkkk  
78: kkkk (K){20} kkkk  
79: kkkk (K){20} kkkk  
80: kkkk (K){20} kkkk  
81: kkkk (K){20} kkkk  
82: kkkk (K){20} kkkk  
83: kkkk (K){20} kkkk  
84: kkkk (K){20} kkkk  
85: kkkk (K){20} kkkk  
86: kkkk (K){20} kkkk  
87: kkkk (K){20} kkkk  
88: kkkk (K){20} kkkk  
89: kkkk (K){20} kkkk  
90: kkkk (K){20} kkkk  
91: kkkk (K){20} kkkk  
92: kkkk (K){20} kkkk  
93: kkkk (K){20} kkkk  
94: kkkk (K){20} kkkk  
95: kkkk (K){20} kkkk  
96: kkkk (K){20} kkkk  
97: kkkk (K){20} kkkk

98: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 99: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 100: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 101: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 102: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 103: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 104: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 105: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 106: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 107: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 108: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 109: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 110: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 111: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 112: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 113: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 114: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 115: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 116: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 117: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 118: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 119: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 120: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 121: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 122: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}

123: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 124: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 125: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 126: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 127: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 128: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 129: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 130: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 131: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 132: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 133: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 134: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 135: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 136: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 137: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 138: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 139: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 140: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 141: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 142: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 143: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 144: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 145: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}  
 146: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
       (K){20}

147: kkkk (k){20} kkkk  
148: kkkk (k){20} kkkk  
149: kkkk (k){20} kkkk  
150: kkkk (k){20} kkkk  
151: kkkk (k){20} kkkk  
152: kkkk (k){20} kkkk  
153: kkkk (k){20} kkkk  
154: kkkk (k){20} kkkk  
155: kkkk (k){20} kkkk  
156: kkkk (k){20} kkkk  
157: kkkk (k){20} kkkk  
158: kkkk (k){20} kkkk  
159: kkkk (k){20} kkkk  
160: kkkk (k){20} kkkk  
161: kkkk (k){20} kkkk  
162: kkkk (k){20} kkkk  
163: kkkk (k){20} kkkk  
164: kkkk (k){20} kkkk  
165: kkkk (k){20} kkkk  
166: kkkk (k){20} kkkk  
167: kkkk (k){20} kkkk  
168: kkkk (k){20} kkkk  
169: kkkk (k){20} kkkk  
170: kkkk (k){20} kkkk

171: kkkk kkkkkkkkkkkkkkkkk kkkk  
172: kkkk (k){20} kkkk  
173: kkkk (k){20} kkkk  
174: kkkk (k){20} kkkk  
175: kkkk (k){20} kkkk  
176: kkkk (k){20} kkkk  
177: kkkk (k){20} kkkk  
178: kkkk (k){20} kkkk  
179: kkkk (k){20} kkkk  
180: kkkk (k){20} kkkk  
181: kkkk (k){20} kkkk  
182: kkkk (k){20} kkkk  
183: kkkk (k){20} kkkk  
184: kkkk (k){20} kkkk  
185: kkkk (k){20} kkkk  
186: kkkk (k){20} kkkk  
187: kkkk (k){20} kkkk  
188: kkkk (k){20} kkkk  
189: kkkk (k){20} kkkk  
190: kkkk (k){20} kkkk  
191: kkkk (k){20} kkkk  
192: kkkk (k){20} kkkk  
193: kkkk (k){20} kkkk  
194: kkkk (k){20} kkkk  
195: kkkk (k){20} kkkk



196: kkkk (k){20} kkkk

197: kkkk (k){20} kkkk

198: kkkk (k){20} kkkk

199: kkkk (k){20} kkkk

200: kkkk (k){20} kkkk

201: kkkk (k){20} kkkk

202: kkkk (k){20} kkkk

203: kkkk (k){20} kkkk

204: kkkk (k){20} kkkk

205: kkkk (k){20} kkkk

206: kkkk (k){20} kkkk

207: kkkk (k){20} kkkk

208: kkkk (k){20} kkkk

209: kkkk (k){20} kkkk

210: kkkk (k){20} kkkk

211: kkkk (k){20} kkkk

212: kkkk (k){20} kkkk

213: kkkk (k){20} kkkk

214: kkkk (k){20} kkkk

215: kkkk (k){20} kkkk

216: kkkk (k){20} kkkk

217: kkkk (k){20} kkkk

218: kkkk (k){20} kkkk

219: kkkk (k){20} kkkk

220: kkkk (k){20} kkkk

221: kkkk (k){20} kkkk

222: kkkk (k){20} kkkk

223: kkkk (k){20} kkkk

224: kkkk (k){20} kkkk

225: kkkk (k){20} kkkk

226: kkkk (k){20} kkkk

227: kkkk (k){20} kkkk

228: kkkk (k){20} kkkk

229: kkkk (k){20} kkkk

230: kkkk (k){20} kkkk

231: kkkk (k){20} kkkk

232: kkkk (k){20} kkkk

233: kkkk (k){20} kkkk

234: kkkk (k){20} kkkk

235: kkkk (k){20} kkkk

236: kkkk (k){20} kkkk

237: kkkk (k){20} kkkk

238: kkkk (k){20} kkkk

239: kkkk (k){20} kkkk

240: kkkk (k){20} kkkk

241: kkkk (k){20} kkkk

242: kkkk (k){20} kkkk

243: kkkk (k){20} kkkk

244: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 245: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 246: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 247: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 248: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 249: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 250: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 251: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 252: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 253: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 254: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 255: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 256: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 257: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 258: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 259: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 260: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 261: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 262: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 263: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 264: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 265: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 266: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 267: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 268: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}

269: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 270: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 271: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 272: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 273: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 274: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 275: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 276: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 277: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 278: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 279: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 280: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 281: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 282: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 283: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 284: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 285: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 286: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 287: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 288: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 289: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 290: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 291: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}  
 292: kkkk kkkkkkkkkkkkkkkkkkk kkkk  
 (K){20}

293: kkkk (K){20} kkkk  
294: kkkk (K){20} kkkk  
295: kkkk (K){20} kkkk  
296: kkkk (K){20} kkkk  
297: kkkk (K){20} kkkk  
298: kkkk (K){20} kkkk  
299: kkkk (K){20} kkkk  
300: kkkk (K){20} kkkk  
301: kkkk (K){20} kkkk  
302: kkkk (K){20} kkkk  
303: kkkk (K){20} kkkk  
304: kkkk (K){20} kkkk  
305: kkkk (K){20} kkkk  
306: kkkk (K){20} kkkk  
307: kkkk (K){20} kkkk  
308: kkkk (K){20} kkkk  
309: kkkk (K){20} kkkk  
310: kkkk (K){20} kkkk  
311: kkkk (K){20} kkkk  
312: kkkk (K){20} kkkk  
313: kkkk (K){20} kkkk  
314: kkkk (K){20} kkkk  
315: kkkk (K){20} kkkk  
316: kkkk (K){20} kkkk

317: kkkk kkkkkkkkkkkkkkkkk kkkk  
318: kkkk (K){20} kkkk  
319: kkkk (K){20} kkkk  
320: kkkk (K){20} kkkk  
321: kkkk (K){20} kkkk  
322: kkkk (K){20} kkkk  
323: kkkk (K){20} kkkk  
324: kkkk (K){20} kkkk  
325: kkkk (K){20} kkkk  
326: kkkk (K){20} kkkk  
327: kkkk (K){20} kkkk  
328: kkkk (K){20} kkkk  
329: kkkk (K){20} kkkk  
330: kkkk (K){20} kkkk  
331: kkkk (K){20} kkkk  
332: kkkk (K){20} kkkk  
333: kkkk (K){20} kkkk  
334: kkkk (K){20} kkkk  
335: kkkk (K){20} kkkk  
336: kkkk (K){20} kkkk  
337: kkkk (K){20} kkkk  
338: kkkk (K){20} kkkk  
339: kkkk (K){20} kkkk  
340: kkkk (K){20} kkkk  
341: kkkk (K){20} kkkk

342: kkkk (K){20} kkkk  
343: kkkk (K){20} kkkk  
344: kkkk (K){20} kkkk  
345: kkkk (K){20} kkkk  
346: kkkk (K){20} kkkk  
347: kkkk (K){20} kkkk  
348: kkkk (K){20} kkkk  
349: kkkk (K){20} kkkk  
350: kkkk (K){20} kkkk  
351: kkkk (K){20} kkkk  
352: kkkk (K){20} kkkk  
353: kkkk (K){20} kkkk  
354: kkkk (K){20} kkkk  
355: kkkk (K){20} kkkk  
356: kkkk (K){20} kkkk  
357: kkkk (K){20} kkkk  
358: kkkk (K){20} kkkk  
359: kkkk (K){20} kkkk  
360: kkkk (K){20} kkkk  
361: kkkk (K){20} kkkk  
362: kkkk (K){20} kkkk  
363: kkkk (K){20} kkkk  
364: kkkk (K){20} kkkk  
365: kkkk (K){20} kkkk

366: kkkk (K){20} kkkk  
367: kkkk (K){20} kkkk  
368: kkkk (K){20} kkkk  
369: kkkk (K){20} kkkk  
370: kkkk (K){20} kkkk  
371: kkkk (K){20} kkkk  
372: kkkk (K){20} kkkk  
373: kkkk (K){20} kkkk  
374: kkkk (K){20} kkkk  
375: kkkk (K){20} kkkk  
376: kkkk (K){20} kkkk  
377: kkkk (K){20} kkkk  
378: kkkk (K){20} kkkk  
379: kkkk (K){20} kkkk  
380: kkkk (K){20} kkkk  
381: kkkk (K){20} kkkk  
382: kkkk (K){20} kkkk  
383: kkkk (K){20} kkkk  
384: kkkk (K){20} kkkk  
385: kkkk (K){20} kkkk  
386: kkkk (K){20} kkkk  
387: kkkk (K){20} kkkk  
388: kkkk (K){20} kkkk  
389: kkkk (K){20} kkkk

390: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 391: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 392: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 393: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 394: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 395: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 396: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 397: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 398: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 399: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 400: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 401: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 402: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 403: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 404: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 405: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 406: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 407: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 408: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 409: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 410: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 411: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 412: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 413: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (K)120  
 414: KKKKK KKKKKKKKKKKKKKKKKKK K

1  
 415: KKKKK KKKKKKKKKKKKKKKKKKK (K)120  
 AAW65939 ck: 1569 len: 40 i Aaw65939 Polylysine peptide NBC32. 11/1:  
 1 (R,K)120  
 2: T KKKKKKKKKKKKKKKKKKKKK (K)120  
 3: TK KKKKKKKKKKKKKKKKKKKKK (K)120  
 4: TKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 5: TKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 6: TKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 7: KKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 8: KKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 9: KKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 10: KKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 11: KKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 12: KKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 13: KKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 14: KKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 15: KKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 16: KKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 17: KKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 18: KKKKK KKKKKKKKKKKKKKKKKKKKK (K)120  
 AAW48808 ck: 9658 len: 56 i Aaw48808 Homo sapiens clone CG109\_1 pro  
 1 (R,K)120  
 34: EEFRE KKKKKKKKKKKKKKKKKKKKK (K)120  
 35: EPREK KKKKKKKKKKKKKKKKKKKKK (K)120  
 36: FREKK KKKKKKKKKKKKKKKKKKKKK (K)120

37: REKKK KKKKKKKKKKKKKKKKKKKKK

AAW45801 ck: 9500 len: 39 ! Aaw45801 One chain of a bombesin dimer. 6/1

```

(R,K){20
(K){20
C KKKKKKKKKKKKKKKKKKK XGGGQ
2:

```

AAW45800 ck: 9056 len: 39 ! Aaw45800 One chain of an alpha-melanocyte s

```

(R,K){20}
(K){20}
18: VGGGX KKKKKKKKKKKKKKKKKKKK CX

```

AAW45802 ck: 7664 len: 35 ! Aaw45802 One chain of an alpha-MSH receptor

```

(R,K){20}
(K){20}
14: VGGGX KKKKKKKKKKKKKKKKKKKK CX

```

AAV43246 ck: 9752 len: 32 ; Aay43246 Cell-surface molecule binding pep

```

12: SGSGS      (R,K){20}
              (K){20}
              KKKKKKKKKKKKKKKKKKKKK K

```

(K){20}

13: GSGSK KKKKKKKKKKKKKKKKKKK

AA10/213 CK: 2211 len: 40 ! Aay0/213 Peptide transfection vector #1. 7/

```

(R,K){20}
(K){20}
21: YEDES KKKKKKKKKKKKKKKKKKKKK

```

AM112950 CR: 3621 len: 62 1 AaY12950 Amino acid sequence of a human seq

```

(R,K){20}
(K){20}
40: QEQAS KKKKKKKKKKKKKKKKKKK KKK

```

```

      (K){20}
41: FQASK KKKKKKKKKKKKKKKKKKK KKK

```

```

42: QASKK KKKKKKKKKKKKKKKKKKK (K){20}
      (K){20}

```

43: ASKKK KKKKKKKKKKKKKKKKKKK

AAB59105	ck: 8456	len: 27	! Aab59105 Breast and ovarian cancer associat
(R,K){20}			
(K){20}			

6: NSAXX KKKKKKKKKKKKKKKKKKKKK (K){20}

```

7: SXXXX KKKKKKKKKKKKKKKKKKKKK K
      (K){20}
8: AXXXX KKKKKKKKKKKKKKKKKKKKK K

```

AAB53249 ck: 4945 len: 59

AA05243	CA:4943	len: 39	AA053249	Human colon cancer antigen protein
(R,K) [20]				

```

      (K){20}
27: ALLSL KKKKKKKKKKKKKKKKKKK KKKKK

```

```

28: LSLK      (K){20}
      KKKKKKKKKKKKKKKKKKKKKKKKKKK

```

```

                (K){20}
29: LSLKK KKKKKKKKKKKKKKKKKKKKK KKKKK

```

```

30: SLKKK KKKKKKKKKKKKKKKKKKK KKNKG
      (K) {20}
      (K) {20}

```

31: LKKKK KKKKKKKKKKKKKKKKKKK KNYCG  
(K) {20}

32: KKKKK KKKKKKKKKKKKKKKKKKKKKK NXXGG

(R, K) [20]  
(K) [20]

```

1:  XXXXXXXXXXXXXXXXXXXX XXXXX
2:  (K){20}
   K XXXXXXXXXXXXXXXXXXXX XXXXX

```

```

      (K) (20)
3 : KK KKKKKKKKKKKKKKKKKKKKK KKKKG

```

```

4:      KKK KKKKKKKKKKKKKKKKKKKKK KXXGG
          (K){20}

```

```

5:      (K){20
      KKKK KKKKKKKKKKKKKKKKKKK KXGGR

```

6: KKKK (K)120j KKKKKKKKKKKKKKKKKKK XGGRF

AAB53800	ck: 296	len: 69	! Aab53800 Human colon cancer antigen prot
	(R,K)[20]		

```
(K){20}
20: NVDC E KKKKKKKKKKKKKKKKKKKKK KK
```

```
21: VDCEK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
```

```

22: DCEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) (20)
23: CEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

```

24: EKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKK

25: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK

26: (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK

```

27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
      (K){20}

```

28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

29: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

```

30: kkkk (K){20} kkkk
31: kkkk (K){20} kkkk
32: kkkk (K){20} kkkk
33: kkkk (K){20} kkkk
34: kkkk (K){20} kkkk
35: kkkk (K){20} kkkk
36: kkkk (K){20} kkkk
37: kkkk (K){20} kkkk
38: kkkk (K){20} kkkk
39: kkkk (K){20} kkkk
40: kkkk (K){20} kkkk

AAB53806 ck: 8373 len: 64 i Aab53806 Human colon cancer antigen protein
(R,K){20}
25: IDCDs kkkk (K){20} kkkk
26: DCDs kkkk (K){20} kkkk
27: CDSk kkkk (K){20} kkkk
28: DSKk kkkk (K){20} kkkk
29: SKKk kkkk (K){20} kkkk
30: kkkk (K){20} kkkk

AAB53977 ck: 7756 len: 75 i Aab53977 Human colon cancer antigen protein
(R,K){20}
24: QPCLL kkkk (K){20} kkkk
25: FCLLk kkkk (K){20} kkkk
26: CLLk kkkk (K){20} kkkk
27: LLKk kkkk (K){20} kkkk
```

```

28: LKKk kkkk (K){20} kkkk
29: kkkk (K){20} kkkk
30: kkkk (K){20} kkkk
31: kkkk (K){20} kkkk
32: kkkk (K){20} kkkk
33: kkkk (K){20} kkkk
34: kkkk (K){20} kkkk
35: kkkk (K){20} kkkk
36: kkkk (K){20} kkkk
37: kkkk (K){20} kkkk
38: kkkk (K){20} kkkk
39: kkkk (K){20} kkkk
40: kkkk (K){20} kkkk
41: kkkk (K){20} kkkk
42: kkkk (K){20} kkkk
43: kkkk (K){20} kkkk
44: kkkk (K){20} kkkk
45: kkkk (K){20} kkkk
46: kkkk (K){20} kkkk
47: kkkk (K){20} kkkk
48: kkkk (K){20} kkkk
49: kkkk (K){20} kkkk

AAB53980 ck: 881 len: 45 i Aab53980 Human colon cancer antigen pro
(R,K){20}
8: ICSFQ kkkk (K){20} kkkk
```

9: CSFOK KKKKKKKKKKKKKKKKKKK KKKXG  
(K){20}  
10: SFQKK KKKKKKKKKKKKKKKKKKK KXXKG  
(K){20}  
11: FOQKK KKKKKKKKKKKKKKKKKKK KXGGR  
(K){20}  
12: OQKKK KKKKKKKKKKKKKKKKKKK XGGRF  
(K){20}  
AAB54314 ck: 6235 len: 55 i Aab54314 Human pancreatic cancer antigen pr  
(R,K){20}  
24: FPYXX KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: FXXKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: XXXKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
27: XXXKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
28: XKKKK KKKKKKKKKKKKKKKKKKK KKKKG  
(K){20}  
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKG  
(K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKKK KKGGR  
(K){20}  
31: KKKKK KKKKKKKKKKKKKKKKKKK KGGRF  
(K){20}  
32: KKKKK KKKKKKKKKKKKKKKKKKK GGGRF  
(K){20}  
AAB56121 ck: 5941 len: 125 i Aab56121 Human secreted protein sequence en  
(R,K){20}  
87: ILKSQ KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
88: LKSQK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
89: KSQKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
90: SQKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
91: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
92: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
93: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
94: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
95: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

1  
96: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
97: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
98: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
99: KKKKK KKKKKKKKKKKKKKKKKKK KKKKG  
(K){20}  
100: KKKKK KKKKKKKKKKKKKKKKKKK KKKKG  
(K){20}  
101: KKKKK KKKKKKKKKKKKKKKKKKK KKGGR  
(K){20}  
102: KKKKK KKKKKKKKKKKKKKKKKKK KGGR  
(K){20}  
103: KKKKK KKKKKKKKKKKKKKKKKKK GGR  
(K){20}  
AAB27956 ck: 6732 len: 139 i Aab27956 Human secreted protein SEQ ID N  
(R,K){20}  
91: HFRPG KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
92: FRPGK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
93: RPKGK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
94: PGKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
95: GKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
96: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
97: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
98: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
99: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
100: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
101: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
102: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
103: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
104: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
105: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}



```
106: KKKK (K){20} KKKK
107: KKKK (K){20} KKKK
108: KKKK (K){20} KKKK
109: KKKK (K){20} KKKK
110: KKKK (K){20} KKKK
111: KKKK (K){20} KKKK
112: KKKK (K){20} KKKK
113: KKKK (K){20} KKKK
114: KKKK (K){20} KKKK
115: KKKK (K){20} KKKK
116: KKKK (K){20} KKKK
117: KKKK (K){20} KKKK
118: KKKK (K){20} KKKK
119: KKKK (K){20} K
120: KKKK (K){20} KKKK

AAB39140 ck: 9531 len: 66 ! Aab39140 Human secreted protein #48. 2/2001
(R,K){20}
41: MMTVX KKKKKKKKKKKKKKKKK KKKK
42: WTVXK KKKKKKKKKKKKKKKKK KKKK
43: TVXKK KKKKKKKKKKKKKKKKK KKKK
44: VXXKK KKKKKKKKKKKKKKKKK KKK
45: XKKKK KKKKKKKKKKKKKKKKK KKK
46: KKKKK (K){20} KKKKKKKKKKKKKKKKK K
47: KKKKK (K){20} KKKKKKKKKKKKKKKKK K

AAB41457 ck: 4536 len: 168 ! Aab41457 Human ORFX ORF1221 polypeptide seq
```

```
144: KQOKP (R,K){20} KKKKKKKKKKKKKKKKK KKKK
145: QOKPK (K){20} KKKKKKKKKKKKKKKKK KKKK
146: QKPKK (K){20} KKKKKKKKKKKKKKKKK KKK
147: KPKKK (K){20} KKKKKKKKKKKKKKKKK KKK
148: PKKKK (K){20} KKKKKKKKKKKKKKKKK K
149: KKKKK (K){20} KKKKKKKKKKKKKKKKK KKKK

AAB42786 ck: 4959 len: 102 ! Aab42786 Human ORFX ORF2550 polypeptide
(R,K){20}
81: SDVLQ KKKKKKKKKKKKKKKKK KKK
82: DVLQK KKKKKKKKKKKKKKKKK K
83: VLQKK KKKKKKKKKKKKKKKKK KKKK

AAB42889 ck: 6432 len: 62 ! Aab42889 Human ORFX ORF2653 polypeptide
(R,K){20}
40: SHRCL KKKKKKKKKKKKKKKKK KKK
41: HRCLK KKKKKKKKKKKKKKKKK KKK
42: RCLKK KKKKKKKKKKKKKKKKK K
43: CLKKK (K){20} KKKKKKKKKKKKKKKKK KKKK

AAB43641 ck: 5216 len: 133 ! Aab43641 Human cancer associated protel
(R,K){20}
114: GHEQS KKKKKKKKKKKKKKKKK KKKK

AAB43835 ck: 4025 len: 223 ! Aab43835 Human cancer associated protel
(R,K){20}
196: NILEW KKKKKKKKKKKKKKKKK KXGGA

AAB44188 ck: 4991 len: 43 ! Aab44188 Human cancer associated protel
(R,K){20}
2: E KKKKKKKKKKKKKKKKK KKKKK
3: EK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
```



5: YKAK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: YKAKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KAKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: AKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AAV9495 ck: 4361 len: 59 i Aay98495 Nuclear ligand used in nucleic ac1  
(R,K){20}  
18: APYKA KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: PYKAK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: YKAKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

21: KAKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: AKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
27: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
28: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
31: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
32: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
33: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
34: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
35: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
36: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
37: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AAV98497 ck: 4925 len: 100 i Aay98497 Peptide #8 used in nucleic ac1  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

6: kkkk (K){20} kkkk  
7: kkkk (K){20} kkkk  
8: kkkk (K){20} kkkk  
9: kkkk (K){20} kkkk  
10: kkkk (K){20} kkkk  
11: kkkk (K){20} kkkk  
12: kkkk (K){20} kkkk  
13: kkkk (K){20} kkkk  
14: kkkk (K){20} kkkk  
15: kkkk (K){20} kkkk  
16: kkkk (K){20} kkkk  
17: kkkk (K){20} kkkk  
18: kkkk (K){20} kkkk  
19: kkkk (K){20} kkkk  
20: kkkk (K){20} kkkk  
21: kkkk (K){20} kkkk  
22: kkkk (K){20} kkkk  
23: kkkk (K){20} kkkk  
24: kkkk (K){20} kkkk  
25: kkkk (K){20} kkkk  
26: kkkk (K){20} kkkk  
27: kkkk (K){20} kkkk  
28: kkkk (K){20} kkkk  
29: kkkk (K){20} kkkk

30: kkkk (K){20} kkkk  
31: kkkk (K){20} kkkk  
32: kkkk (K){20} kkkk  
33: kkkk (K){20} kkkk  
34: kkkk (K){20} kkkk  
35: kkkk (K){20} kkkk  
36: kkkk (K){20} kkkk  
37: kkkk (K){20} kkkk  
38: kkkk (K){20} kkkk  
39: kkkk (K){20} kkkk  
40: kkkk (K){20} kkkk  
41: kkkk (K){20} kkkk  
42: kkkk (K){20} kkkk  
43: kkkk (K){20} kkkk  
44: kkkk (K){20} kkkk  
45: kkkk (K){20} kkkk  
46: kkkk (K){20} kkkk  
47: kkkk (K){20} kkkk  
48: kkkk (K){20} kkkk  
49: kkkk (K){20} kkkk  
50: kkkk (K){20} kkkk  
51: kkkk (K){20} kkkk  
52: kkkk (K){20} kkkk  
53: kkkk (K){20} kkkk

54: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
55: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
56: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
57: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
58: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
59: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
60: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
61: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
62: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
63: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
64: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
65: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
66: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
67: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
68: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
69: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
70: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
71: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
72: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
73: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
74: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
75: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
76: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
77: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
78: KKKK KKKKKKKKKKKKKKKKK KKK  
(K){20}

1

79: KKKK KKKKKKKKKKKKKKKKK K  
(K){20}  
80: KKKK KKKKKKKKKKKKKKKKK K  
(K){20}  
81: KKKK KKKKKKKKKKKKKKKKK  
AAV81805 ck: 1885 len: 351 i Aay81805 murine mahogany protein sequen  
(R,K){20}  
3: EF KKKKKKKKKKKKKKKKK KKKK  
(R,K){20}  
4: EFR KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
5: EFRK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
6: -EPRK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
7: FRKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
8: RKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
9: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
10: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
11: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
12: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
13: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
14: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
15: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
16: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
17: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
18: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
19: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
20: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
21: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
22: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

23: KKKKK (K){20} KKKKK  
24: KKKKK (K){20} KKKKK  
25: KKKKK (K){20} KKKKK  
26: KKKKK (K){20} KKKKK  
27: KKKKK (K){20} KKKKK

AAY56902 ck: 4875 len: 30 i Aay56902 (Lys)30 protein sequence. 4/2000

1: KKKKK (R,K){20} KKKKK  
2: K (K){20} KKKKK  
3: KK (K){20} KKKKK  
4: KKK (K){20} KKKKK  
5: KKKKK (K){20} KKKKK  
6: KKKKK (K){20} KKKKK  
7: KKKKK (K){20} KKKKK  
8: KKKKK (K){20} KKKKK  
9: KKKKK (K){20} KKKKK  
10: KKKKK (K){20} KKKKK  
11: KKKKK (K){20} KKKKK

AAY56903 ck: 5075 len: 434 i Aay56903 (Lys)434 protein sequence. 4/2000

1: KKKKK (R,K){20} KKKKK  
2: K (K){20} KKKKK  
3: KK (K){20} KKKKK  
4: KKK (K){20} KKKKK  
5: KKKKK (K){20} KKKKK

6: KKKKK KKKKK  
7: KKKKK (K){20} KKKKK  
8: KKKKK (K){20} KKKKK  
9: KKKKK (K){20} KKKKK  
10: KKKKK (K){20} KKKKK  
11: KKKKK (K){20} KKKKK  
12: KKKKK (K){20} KKKKK  
13: KKKKK (K){20} KKKKK  
14: KKKKK (K){20} KKKKK  
15: KKKKK (K){20} KKKKK  
16: KKKKK (K){20} KKKKK  
17: KKKKK (K){20} KKKKK  
18: KKKKK (K){20} KKKKK  
19: KKKKK (K){20} KKKKK  
20: KKKKK (K){20} KKKKK  
21: KKKKK (K){20} KKKKK  
22: KKKKK (K){20} KKKKK  
23: KKKKK (K){20} KKKKK  
24: KKKKK (K){20} KKKKK  
25: KKKKK (K){20} KKKKK  
26: KKKKK (K){20} KKKKK  
27: KKKKK (K){20} KKKKK  
28: KKKKK (K){20} KKKKK  
29: KKKKK (K){20} KKKKK  
30: KKKKK (K){20} KKKKK

31: KKKK (K){20} KKKK  
32: KKKK (K){20} KKKK  
33: KKKK (K){20} KKKK  
34: KKKK (K){20} KKKK  
35: KKKK (K){20} KKKK  
36: KKKK (K){20} KKKK  
37: KKKK (K){20} KKKK  
38: KKKK (K){20} KKKK  
39: KKKK (K){20} KKKK  
40: KKKK (K){20} KKKK  
41: KKKK (K){20} KKKK  
42: KKKK (K){20} KKKK  
43: KKKK (K){20} KKKK  
44: KKKK (K){20} KKKK  
45: KKKK (K){20} KKKK  
46: KKKK (K){20} KKKK  
47: KKKK (K){20} KKKK  
48: KKKK (K){20} KKKK  
49: KKKK (K){20} KKKK  
50: KKKK (K){20} KKKK  
51: KKKK (K){20} KKKK  
52: KKKK (K){20} KKKK  
53: KKKK (K){20} KKKK  
54: KKKK (K){20} KKKK

55: KKKK (K){20} KKKK  
56: KKKK (K){20} KKKK  
57: KKKK (K){20} KKKK  
58: KKKK (K){20} KKKK  
59: KKKK (K){20} KKKK  
60: KKKK (K){20} KKKK  
61: KKKK (K){20} KKKK  
62: KKKK (K){20} KKKK  
63: KKKK (K){20} KKKK  
64: KKKK (K){20} KKKK  
65: KKKK (K){20} KKKK  
66: KKKK (K){20} KKKK  
67: KKKK (K){20} KKKK  
68: KKKK (K){20} KKKK  
69: KKKK (K){20} KKKK  
70: KKKK (K){20} KKKK  
71: KKKK (K){20} KKKK  
72: KKKK (K){20} KKKK  
73: KKKK (K){20} KKKK  
74: KKKK (K){20} KKKK  
75: KKKK (K){20} KKKK  
76: KKKK (K){20} KKKK  
77: KKKK (K){20} KKKK  
78: KKKK (K){20} KKKK

79: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 80: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 81: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 82: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 83: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 84: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 85: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 86: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 87: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 88: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 89: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 90: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 91: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 92: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 93: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 94: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 95: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 96: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 97: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 98: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 99: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 100: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 101: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 102: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 103: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}

104: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 105: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 106: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 107: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 108: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 109: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 110: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 111: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 112: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 113: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 114: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 115: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 116: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 117: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 118: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 119: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 120: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 121: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 122: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 123: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 124: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 125: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 126: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}  
 127: kkkk kkkkkkkkkkkkkkkkk kkkk  
 (k){20}



128: kkkk (K) (20) kkkk  
 129: kkkk (K) (20) kkkk  
 130: kkkk (K) (20) kkkk  
 131: kkkk (K) (20) kkkk  
 132: kkkk (K) (20) kkkk  
 133: kkkk (K) (20) kkkk  
 134: kkkk (K) (20) kkkk  
 135: kkkk (K) (20) kkkk  
 136: kkkk (K) (20) kkkk  
 137: kkkk (K) (20) kkkk  
 138: kkkk (K) (20) kkkk  
 139: kkkk (K) (20) kkkk  
 140: kkkk (K) (20) kkkk  
 141: kkkk (K) (20) kkkk  
 142: kkkk (K) (20) kkkk  
 143: kkkk (K) (20) kkkk  
 144: kkkk (K) (20) kkkk  
 145: kkkk (K) (20) kkkk  
 146: kkkk (K) (20) kkkk  
 147: kkkk (K) (20) kkkk  
 148: kkkk (K) (20) kkkk  
 149: kkkk (K) (20) kkkk  
 150: kkkk (K) (20) kkkk  
 151: kkkk (K) (20) kkkk

152: kkkk kkkk kkkk kkkk kkkk  
 153: kkkk (K) (20) kkkk  
 154: kkkk (K) (20) kkkk  
 155: kkkk (K) (20) kkkk  
 156: kkkk (K) (20) kkkk  
 157: kkkk (K) (20) kkkk  
 158: kkkk (K) (20) kkkk  
 159: kkkk (K) (20) kkkk  
 160: kkkk (K) (20) kkkk  
 161: kkkk (K) (20) kkkk  
 162: kkkk (K) (20) kkkk  
 163: kkkk (K) (20) kkkk  
 164: kkkk (K) (20) kkkk  
 165: kkkk (K) (20) kkkk  
 166: kkkk (K) (20) kkkk  
 167: kkkk (K) (20) kkkk  
 168: kkkk (K) (20) kkkk  
 169: kkkk (K) (20) kkkk  
 170: kkkk (K) (20) kkkk  
 171: kkkk (K) (20) kkkk  
 172: kkkk (K) (20) kkkk  
 173: kkkk (K) (20) kkkk  
 174: kkkk (K) (20) kkkk  
 175: kkkk (K) (20) kkkk  
 176: kkkk (K) (20) kkkk

177: kkkk (k){20} kkkk  
178: kkkk (k){20} kkkk  
179: kkkk (k){20} kkkk  
180: kkkk (k){20} kkkk  
181: kkkk (k){20} kkkk  
182: kkkk (k){20} kkkk  
183: kkkk (k){20} kkkk  
184: kkkk (k){20} kkkk  
185: kkkk (k){20} kkkk  
186: kkkk (k){20} kkkk  
187: kkkk (k){20} kkkk  
188: kkkk (k){20} kkkk  
189: kkkk (k){20} kkkk  
190: kkkk (k){20} kkkk  
191: kkkk (k){20} kkkk  
192: kkkk (k){20} kkkk  
193: kkkk (k){20} kkkk  
194: kkkk (k){20} kkkk  
195: kkkk (k){20} kkkk  
196: kkkk (k){20} kkkk  
197: kkkk (k){20} kkkk  
198: kkkk (k){20} kkkk  
199: kkkk (k){20} kkkk  
200: kkkk (k){20} kkkk

201: kkkk (k){20} kkkk  
202: kkkk (k){20} kkkk  
203: kkkk (k){20} kkkk  
204: kkkk (k){20} kkkk  
205: kkkk (k){20} kkkk  
206: kkkk (k){20} kkkk  
207: kkkk (k){20} kkkk  
208: kkkk (k){20} kkkk  
209: kkkk (k){20} kkkk  
210: kkkk (k){20} kkkk  
211: kkkk (k){20} kkkk  
212: kkkk (k){20} kkkk  
213: kkkk (k){20} kkkk  
214: kkkk (k){20} kkkk  
215: kkkk (k){20} kkkk  
216: kkkk (k){20} kkkk  
217: kkkk (k){20} kkkk  
218: kkkk (k){20} kkkk  
219: kkkk (k){20} kkkk  
220: kkkk (k){20} kkkk  
221: kkkk (k){20} kkkk  
222: kkkk (k){20} kkkk  
223: kkkk (k){20} kkkk  
224: kkkk (k){20} kkkk



274: kkkk (K){20} kkkk  
275: kkkk (K){20} kkkk  
276: kkkk (K){20} kkkk  
277: kkkk (K){20} kkkk  
278: kkkk (K){20} kkkk  
279: kkkk (K){20} kkkk  
280: kkkk (K){20} kkkk  
281: kkkk (K){20} kkkk  
282: kkkk (K){20} kkkk  
283: kkkk (K){20} kkkk  
284: kkkk (K){20} kkkk  
285: kkkk (K){20} kkkk  
286: kkkk (K){20} kkkk  
287: kkkk (K){20} kkkk  
288: kkkk (K){20} kkkk  
289: kkkk (K){20} kkkk  
290: kkkk (K){20} kkkk  
291: kkkk (K){20} kkkk  
292: kkkk (K){20} kkkk  
293: kkkk (K){20} kkkk  
294: kkkk (K){20} kkkk  
295: kkkk (K){20} kkkk  
296: kkkk (K){20} kkkk  
297: kkkk (K){20} kkkk

298: kkkk (K){20} kkkk  
299: kkkk (K){20} kkkk  
300: kkkk (K){20} kkkk  
301: kkkk (K){20} kkkk  
302: kkkk (K){20} kkkk  
303: kkkk (K){20} kkkk  
304: kkkk (K){20} kkkk  
305: kkkk (K){20} kkkk  
306: kkkk (K){20} kkkk  
307: kkkk (K){20} kkkk  
308: kkkk (K){20} kkkk  
309: kkkk (K){20} kkkk  
310: kkkk (K){20} kkkk  
311: kkkk (K){20} kkkk  
312: kkkk (K){20} kkkk  
313: kkkk (K){20} kkkk  
314: kkkk (K){20} kkkk  
315: kkkk (K){20} kkkk  
316: kkkk (K){20} kkkk  
317: kkkk (K){20} kkkk  
318: kkkk (K){20} kkkk  
319: kkkk (K){20} kkkk  
320: kkkk (K){20} kkkk  
321: kkkk (K){20} kkkk  
322: kkkk (K){20} kkkk

323: KKKK (K){20} KKKK  
324: KKKK (K){20} KKKK  
325: KKKK (K){20} KKKK  
326: KKKK (K){20} KKKK  
327: KKKK (K){20} KKKK  
328: KKKK (K){20} KKKK  
329: KKKK (K){20} KKKK  
330: KKKK (K){20} KKKK  
331: KKKK (K){20} KKKK  
332: KKKK (K){20} KKKK  
333: KKKK (K){20} KKKK  
334: KKKK (K){20} KKKK  
335: KKKK (K){20} KKKK  
336: KKKK (K){20} KKKK  
337: KKKK (K){20} KKKK  
338: KKKK (K){20} KKKK  
339: KKKK (K){20} KKKK  
340: KKKK (K){20} KKKK  
341: KKKK (K){20} KKKK  
342: KKKK (K){20} KKKK  
343: KKKK (K){20} KKKK  
344: KKKK (K){20} KKKK  
345: KKKK (K){20} KKKK  
346: KKKK (K){20} KKKK

347: KKKK (K){20} KKKK  
348: KKKK (K){20} KKKK  
349: KKKK (K){20} KKKK  
350: KKKK (K){20} KKKK  
351: KKKK (K){20} KKKK  
352: KKKK (K){20} KKKK  
353: KKKK (K){20} KKKK  
354: KKKK (K){20} KKKK  
355: KKKK (K){20} KKKK  
356: KKKK (K){20} KKKK  
357: KKKK (K){20} KKKK  
358: KKKK (K){20} KKKK  
359: KKKK (K){20} KKKK  
360: KKKK (K){20} KKKK  
361: KKKK (K){20} KKKK  
362: KKKK (K){20} KKKK  
363: KKKK (K){20} KKKK  
364: KKKK (K){20} KKKK  
365: KKKK (K){20} KKKK  
366: KKKK (K){20} KKKK  
367: KKKK (K){20} KKKK  
368: KKKK (K){20} KKKK  
369: KKKK (K){20} KKKK  
370: KKKK (K){20} KKKK

371: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
372: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
373: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
374: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
375: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
376: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
377: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
378: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
379: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
380: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
381: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
382: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
383: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
384: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
385: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
386: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
387: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
388: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
389: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
390: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
391: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
392: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
393: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
394: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
395: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}

---

1

396: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
397: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
398: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
399: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
400: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
401: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
402: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
403: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
404: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
405: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
406: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
407: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
408: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
409: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
410: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
411: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
412: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
413: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
414: kkkk kkkkkkkkkkkkkkkkk k  
      (k){20}  
415: kkkk kkkkkkkkkkkkkkkkk kkkk  
      (k){20}

AAy86248 ck: 8783 len: 128 i Aay86248 Human secreted protein HCHPF68,  
      (R,K){20}  
84: ANPP kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
85: NPPK kkkkkkkkkkkkkkkkk kkkk  
      (k){20}  
86: PPPK kkkkkkkkkkkkkkkkk kkkk  
      (k){20}

```
      (K){20}
87: PPKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
88: PKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
89: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
90: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
91: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
92: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
93: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
94: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
95: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
96: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
97: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
98: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
99: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
100: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
101: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
102: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
103: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
104: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
105: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
106: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}

AAV59038 ck: 8137 len: 45  i Aay59038 Peptide used in the construction c
      (R,K){20}
      (K){20}
4: YKA KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
5: YKA KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
6: YKA KKKKKKKKKKKKKKKKKKK KKKK
```

```
      (K){20}
7: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
8: AKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
9: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
10: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
11: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
12: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
13: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
14: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
15: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
16: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
17: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
18: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
19: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
20: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
21: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
22: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
23: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
24: KKKK KKKKKKKKKKKKKKKKKKK KKKK

AAV59040 ck: 4361 len: 59  i Aay59040 Nuclear ligand for transportin
      (R,K){20}
      (K){20}
18: APYKA KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
19: PYKA KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
20: YKA KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
21: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
22: AKKK KKKKKKKKKKKKKKKKKKK KKKK
```

23: KKKK (K){20} KKKK  
24: KKKK (K){20} KKKK  
25: KKKK (K){20} KKKK  
26: KKKK (K){20} KKKK  
27: KKKK (K){20} KKKK  
28: KKKK (K){20} KKKK  
29: KKKK (K){20} KKKK  
30: KKKK (K){20} KKKK  
31: KKKK (K){20} KKKK  
32: KKKK (K){20} KKKK  
33: KKKK (K){20} KKKK  
34: KKKK (K){20} KKKK  
35: KKKK (K){20} KKKK  
36: KKKK (K){20} KKKK  
37: KKKK (K){20} KKKK  
38: KKKK (K){20} KKKK  
AA59042 ck: 4925 len: 100 i Aay59042 Amino acid polymer seq ID NO: 62 C  
1: (R,K){20} KKKK  
2: (K){20} KKKK  
3: (K){20} KKKK  
4: (K){20} KKKK  
5: (K){20} KKKK  
6: (K){20} KKKK  
7: (K){20} KKKK

8: KKKK (K){20} KKKK  
9: KKKK (K){20} KKKK  
10: KKKK (K){20} KKKK  
11: KKKK (K){20} KKKK  
12: KKKK (K){20} KKKK  
13: KKKK (K){20} KKKK  
14: KKKK (K){20} KKKK  
15: KKKK (K){20} KKKK  
16: KKKK (K){20} KKKK  
17: KKKK (K){20} KKKK  
18: KKKK (K){20} KKKK  
19: KKKK (K){20} KKKK  
20: KKKK (K){20} KKKK  
21: KKKK (K){20} KKKK  
22: KKKK (K){20} KKKK  
23: KKKK (K){20} KKKK  
24: KKKK (K){20} KKKK  
25: KKKK (K){20} KKKK  
26: KKKK (K){20} KKKK  
27: KKKK (K){20} KKKK  
28: KKKK (K){20} KKKK  
29: KKKK (K){20} KKKK  
30: KKKK (K){20} KKKK  
31: KKKK (K){20} KKKK



32: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
33: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
34: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
35: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
36: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
37: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
38: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
39: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
40: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
41: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
42: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
43: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
44: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
45: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
46: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
47: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
48: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
49: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
50: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
51: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
52: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
53: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
54: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
55: KKKK (K) [20]  
KKKKKKKKKKKKKKKKKK KKKK  
(K) [20]

56: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
57: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
58: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
59: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
60: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
61: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
62: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
63: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
64: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
65: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
66: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
67: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
68: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
69: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
70: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
71: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
72: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
73: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
74: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
75: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
76: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
77: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
78: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
79: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]  
80: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K) [20]

81: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

ABG00401 ck: 5116 len: 1,074 ! Abg00401 Novel human diagnostic protein #39

(R,K){20}  
(R,K){20}

609: RGSSS KKKRRKKKKKKKKRRR KNKK

(R,K){20}

610: GSSSK KKKRRKKKKKKRRR NRKK

ABG03974 ck: 5341 len: 99 ! Abg03974 Novel human diagnostic protein #39

(R,K){20}

2: M KKKKKKKKKKKKKKKKK

3: MK KKKKKKKKKKKKKKKKK

4: MKK KKKKKKKKKKKKKKKKK

5: MKKK KKKKKKKKKKKKKKKKK NSOI

6: MKKKK KKKKKKKKKKKKKKKKK NSOID

ABG04391 ck: 9047 len: 139 ! Abg04391 Novel human diagnostic protein #43

(R,K){20}

79: EEEEE KKKKKKKKKKKKKKKKK

80: EEEBK KKKKKKKKKKKKKKKKK

81: EEEKK KKKKKKKKKKKKKKKKK

82: EEEKK KKKKKKKKKKKKKKKKK

83: EKKKK KKKKKKKKKKKKKKKKK

84: KKKKK KKKKKKKKKKKKKKKKK

85: KKKKK KKKKKKKKKKKKKKKKK

86: KKKKK KKKKKKKKKKKKKKKKK

87: KKKKK KKKKKKKKKKKKKKKKK

88: KKKKK KKKKKKKKKKKKKKKKK

89: KKKKK KKKKKKKKKKKKKKKKK

90: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

91: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

92: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

93: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

94: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

95: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

96: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

97: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

98: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

99: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

100: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

101: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

102: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

103: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

104: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

105: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

106: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

107: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

108: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

109: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

110: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

111: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

112: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

113: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK

ABG05352 ck: 1276 len: 204 ! Abg05352 Novel human diagnostic protein

(R,K){20}

(K) {20}  
106: EEEEE KKKKKKKKKKKKKKKKKKK KKKK  
(K) {20}  
107: EEEK KKKKKKKKKKKKKKKKKKK KKKK  
(K) {20}  
108: EEEK KKKKKKKKKKKKKKKKKKK KKKK  
(K) {20}  
109: EEKK KKKKKKKKKKKKKKKKKKK KKKK  
(K) {20}  
110: EKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K) {20}  
111: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K) {20}  
112: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(R, K) {20}  
113: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(R, K) {20}  
114: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(R, K) {20}  
115: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(R, K) {20}  
116: KKKK KKKKKKKKKKKKKKKKKKK KKEE  
(R, K) {20}  
117: KKKK KKKKKKKKKKKKKKKKKKK KEEE  
(R, K) {20}  
118: KKKK KKKKKKKKKKKKKKKKKKK EEEK

1  
ABG05367 ck: 6907 len: 808 ! Abg05367 Novel human diagnostic protein #53  
(R, K) {20}  
(R, K) {20}  
219: EEEE RKKKKKKKKKKKKKKKKKK RRRR  
(R, K) {20}  
220: EEEER RKKKKKKKKKKKKKKKKKK RRRR  
(R, K) {20}  
221: EEER KKKKKKKKKKKKKKKKKKK RRRR  
(R, K) {20}  
222: EERR KKKKKKKKKKKKKKKKKKK RRRR  
(R, K) {20}  
223: ERRK KKKKKKKKKKKKKKKKKKK RRRR  
(R, K) {20}  
224: RKKK KKKKKKKKKKKKKKKKKKK RRRR  
(R, K) {20}  
225: RKKK KKKKKKKKKKKKKKKKKKK RRRR  
(R, K) {20}  
226: KKKK KKKKKKKKKKKKKKKKKKK RRRR  
(R, K) {20}  
227: KKKK KKKKKKKKKKKKKKKKKKK RRRR  
(R, K) {20}  
228: KKKK KKKKKKKKKKKKKKKKKKK RRRR

(R, K) {20}  
229: KKKK RRRRRRRRRRRRRRRRRRR KKKK  
(R, K) {20}  
230: KKKK RRRRRRRRRRRRRRRRRRR KKKK  
(R, K) {20}  
231: KKKR RRRRRRRRRRRRRRRRRRR KKKK  
(R, K) {20}  
232: KKKR RRRRRRRRRRRRRRRRRRR KKKK  
(R, K) {20}  
233: KKKR RRRRRRRRRRRRRRRRRRR KKKK  
(R, K) {20}  
234: KKKR RRRRRRRRRRRRRRRRRRR KKKK  
(R, K) {20}  
235: KKKR RRRRRRRRRRRRRRRRRRR KKKK  
(R, K) {20}  
236: KKKR RRRRRRRRRRRRRRRRRRR KKKK  
(R, K) {20}  
237: KKKR RRRRRRRRRRRRRRRRRRR KKKK  
(R, K) {20}  
238: KKKR RRRRRRRRRRRRRRRRRRR KKKK  
(R, K) {20}  
239: KKKR RRRRRRRRRRRRRRRRRRR KKKK  
(R, K) {20}  
240: KKKR RRRRRRRRRRRRRRRRRRR KKKK  
(R, K) {20}

1  
ABG06375 ck: 7807 len: 2,570 ! Abg06375 Novel human diagnostic protein  
(R, K) {20}  
(R, K) {20}  
205: KKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R, K) {20}  
206: KKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R, K) {20}  
207: KKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R, K) {20}  
208: KKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R, K) {20}  
ABG06513 ck: 2934 len: 154 ! Abg06513 Novel human diagnostic protein  
(R, K) {20}  
(R, K) {20}  
49: KKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R, K) {20}  
50: KKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R, K) {20}  
51: KKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R, K) {20}  
52: KKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R, K) {20}  
53: KKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R, K) {20}  
54: KKKK RRRRRRRRRRRRRRRRRRR RRRR

[illegible]

66: KKKKK KKKKKKKKKKKKKKKKKKK KNLLY  
 (R,K){20}  
 67: KKKKK KKKKKKKKKKKKKKKKKKK NLLYR  
 (R,K){20}  
 1  
 ABG10052 ck: 7107 len: 39 i Abg10052 Novel human diagnostic protein  
 (R,K){20}  
 2: Q KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 3: QK KKKKKKKKKKKKKKKKKKK KKKKE  
 (K){20}  
 4: QKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 5: QKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 6: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 7: KKKKK KKKKKKKKKKKKKKKKKKK EKKKK  
 (K){20}  
 1  
 ABG10053 ck: 3274 len: 189 i Abg10053 Novel human diagnostic protein  
 (R,K){20}  
 70: EKEKE KKKKKKKKKKKKKKKKKKK KKKKK  
 (R,K){20}  
 71: KEKEK KKKKKKKKKKKKKKKKKKK KKKKK  
 (R,K){20}  
 72: EKEKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (R,K){20}  
 73: KEKKR KKKKKKKKKKKKKKKKKKK KKKKK  
 (R,K){20}  
 74: EKKRK KKKKKKKKKKKKKKKKKKK KKKKK  
 (R,K){20}  
 75: KKKKK KKKKKKKKKKKKKKKKKKK KKKKE  
 (R,K){20}  
 76: KKKKK KKKKKKKKKKKKKKKKKKK KKKKE  
 (R,K){20}  
 77: RKKKR KKKKKKKKKKKKKKKKKKK KKEEE  
 (K){20}  
 78: KKKRK KKKKKKKKKKKKKKKKKKK KEEEE  
 (K){20}  
 79: KKKKK KKKKKKKKKKKKKKKKKKK EEEEE  
 (K){20}  
 1  
 ABG11241 ck: 3870 len: 121 i Abg11241 Novel human diagnostic protein  
 (R,K){20}  
 44: KKKKE KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 45: KKEEK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}

46: KEEKK (K){20} KKKKKKKKKKKKKKKKKKK KKKXE  
(K){20}  
47: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKXK  
(K){20}  
48: EKKKK KKKKKKKKKKKKKKKKKKKKK KXEXE  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKKKKKK XEXEK

ABG11242 ck: 4493 len: 100 ! Abg11242 Novel human diagnostic protein #11

(R,K){20}  
42: ETPSE KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
43: TPSEK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
44: PSEKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
45: SEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
46: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
50: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
51: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
55: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

1

(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
65: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
66: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
67: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
68: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
69: KKKKK KKKKKKKKKKKKKKKKKKKKK KKNPI  
(K){20}  
70: KKKKK KKKKKKKKKKKKKKKKKKKKK KNPIF  
(K){20}  
71: KKKKK KKKKKKKKKKKKKKKKKKKKK NPIFE  
(K){20}

ABG11245 ck: 2517 len: 85 ! Abg11245 Novel human diagnostic protein

(R,K){20}  
33: EEEEE KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
34: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
35: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
36: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
37: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
39: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
40: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
41: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
42: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
44: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

45: KKKKK (K){20} KKKKK  
46: KKKKK (K){20} KKKKK  
47: KKKKK (K){20} KKKKK  
48: KKKKK (K){20} KKKKK  
49: KKKKK (K){20} KKKKK  
50: KKKKK (K){20} KKKKK  
51: KKKKK (K){20} KKKKK  
52: KKKKK (K){20} KKKKK

ABG11250 ck: 4343 len: 92 ; Abg11250 Novel human diagnostic protein #11

53: EKEKE (R,K){20} KKKKK  
54: KEKEK (K){20} KKKKK  
55: EKEKK (K){20} KKKKK  
56: KEKKK (K){20} KKKKK  
57: EKKKK (R,K){20} KKKKK  
58: KKKKK (R,K){20} KKKKK  
59: KKKKK (R,K){20} KKKKK

ABG11266 ck: 9563 len: 146 ; Abg11266 Novel human diagnostic protein #11

36: KEKRE (R,K){20} KKKKK  
37: EKREK (K){20} KKKKK  
38: KREKK (K){20} KKKKK  
39: REKKK (K){20} KKKKK  
40: EKKKK (K){20} KKKKK  
79: EEEEE (R,K){20} KKKKK  
80: EEEEE (R,K){20} KKKKK

1

81: EEEKK (R,K){20} KKKKK  
82: EEEKK (R,K){20} KKKKK  
83: EKKKK (R,K){20} KKKKK  
84: KKKKK (R,K){20} KKKKK  
85: KKKKK (R,K){20} KKKKK  
86: KKKKK (R,K){20} KKKKK  
87: KKKKK (R,K){20} KKKKK  
88: KKKKK (R,K){20} KKKKK  
89: KKKKK (R,K){20} KKKKK  
90: KKKKK (R,K){20} KKKKK  
91: KKKKK (R,K){20} KKKKK  
92: KKKKK (R,K){20} KKKKK  
93: KKKKK (R,K){20} KKKKK

ABG11277 ck: 8026 len: 1,080 ; Abg11277 Novel human diagnostic protein

709: OKEKE (R,K){20} KKKKK  
710: KEKEK (R,K){20} KKKKK  
711: EKEKK (R,K){20} KKKKK  
712: KEKKK (R,K){20} KKKKK  
713: EKKKK (R,K){20} KKKKK  
714: KKKKK (R,K){20} KKKKK  
715: KKKKK (R,K){20} KKKKK  
716: KKKKK (R,K){20} KKKKK  
717: KKKKK (R,K){20} KKKKK  
718: KKKKK (R,K){20} KKKKK

(R,K){20}  
719: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
720: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
721: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
722: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
723: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
724: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
725: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
726: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
727: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
ABG11732 ck: 2886 len: 56 i Abg11732 Novel human diagnostic protein #11  
(R,K){20}  
8: KRRRG RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
9: RRRGR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
10: RRGRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
11: RGRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
12: GRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
13: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
ABG11734 ck: 4548 len: 142 i Abg11734 Novel human diagnostic protein #11  
(R,K){20}  
109: QMLSV KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
110: MLSVK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
111: LSVKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
112: SVKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
113: VKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
114: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

115: KKKRR KKKKKKKKKKKKKKKKKKK KKKLY  
(K){20}  
116: KKKRR KKKKKKKKKKKKKKKKKKK KKLVE  
(K){20}  
117: KKKRR KKKKKKKKKKKKKKKKKKK KLYPO  
(K){20}  
118: KKKRR KKKKKKKKKKKKKKKKKKK LYFOT  
ABG11738 ck: 3009 len: 567 i Abg11738 Novel human diagnostic protein  
(R,K){20}  
526: EEEEE RRRKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
527: EEEER RKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
528: EEEER KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
529: EEEER KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
530: ERRKK KKKKKKKKKKKKKKKKKKK KKKKE  
(K){20}  
531: RRRKK KKKKKKKKKKKKKKKKKKK KRRER  
(K){20}  
532: RKKKK KKKKKKKKKKKKKKKKKKK RKEER  
(R,K){20}  
533: KKKKK KKKKKKKKKKKKKKKKKKK KERRT  
(R,K){20}  
534: KKKKK KKKKKKKKKKKKKKKKKKK ERRTA  
ABG22512 ck: 8641 len: 856 i Abg22512 Novel human diagnostic protein  
(R,K){20}  
130: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
131: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
132: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
133: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
134: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
135: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
136: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
ABG22638 ck: 4197 len: 896 i Abg22638 Novel human diagnostic protein  
(R,K){20}  
227: GREER RRRRRRRRRRRRRRRRRRR RRRRR

228: REPER (R,K)(20)  
RRRRRRRRRRRRRRRRRRRRRRRR  
229: EPERR (R,K)(20)  
RRRRRRRRRRRRRRRRRRRRRRRGXE  
230: RERRR (R,K)(20)  
RRRRRRRRRRRRRRRRRRRRRRRGXEF  
231: ERRRR (R,K)(20)  
RRRRRRRRRRRRRRRRRRRRRRGXEFL

1  
ABG26213 ck: 6773 len: 735 i Abg26213 Novel human diagnostic protein #26  
(R,K)(20)  
173: RGSSS (R,K)(20)  
KKKKRRKKKKKKKKKKRRR KNKK  
174: GSSSK (R,K)(20)  
KKRRKKKKKKKKRRR NRKK

1  
ABG26488 ck: 523 len: 124 i Abg26488 Novel human diagnostic protein #26  
(R,K)(20)  
91: EEEEE RRRRRKKKKKKKKKKR KKKK  
92: EEEER RRRKKKKKKKKKKRR KKKK  
(R,K)(20)  
93: EEEER RRRKKKKKKKKKKRR KKKK  
(R,K)(20)  
94: EERRR RRRKKKKKKKKKKRR KKKK  
(R,K)(20)  
95: ERRRR RKKKKKKKKKKKKRR KKKK  
(R,K)(20)  
96: RRRRR KKKKKKKKKKKKKRR KKKK  
(R,K)(20)  
97: RRRRR KKKKKKKKKKKKKRR KKKK  
(R,K)(20)  
98: RRRRR KKKKKKKKKKKKKRR KKKK  
(R,K)(20)  
99: RRRRR KKKKKKKKKKKKKRR KKKK  
(R,K)(20)  
100: RRRRR KKKKKKKKKKKKKRR KKKK  
(R,K)(20)  
101: KKKKK KKKKKKKKKKKKKRR KKKK  
(R,K)(20)  
102: KKKKK KKKKKKKKKKKKKRR KKKK  
(R,K)(20)  
103: KKKKK KKKKKKKKKKKKKRR KE  
(R,K)(20)  
104: KKKKK KKKKKKKKKKKKKRR E  
(R,K)(20)  
ABG26489 ck: 9923 len: 120 i Abg26489 Novel human diagnostic protein #26  
(R,K)(20)

(R)(20)  
70: RRRRG RRRRRRRRRRRRRRRRRRRRRRRRR  
(R)(20)  
71: RRRGR RRRRRRRRRRRRRRRRRRRRRRRRR  
(R)(20)  
72: RRGRR RRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K)(20)  
73: RGRRR RRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K)(20)  
74: GRRRR RRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K)(20)  
75: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K)(20)

1  
ABG26490 ck: 4146 len: 96 i Abg26490 Novel human diagnostic protein  
(R,K)(20)  
39: KKGEE RRRRRRRRRRRRRRRRRRRRRRRRR  
(R)(20)  
40: KGEER RRRRRRRRRRRRRRRRRRRRRRRRR  
(R)(20)  
41: GEERR RRRRRRRRRRRRRRRRRRRRRRRRR  
(R)(20)  
42: EERRR RRRRRRRRRRRRRRRRRRRRRRRRR  
(R)(20)  
73: GRGGR RRRRRRRRRRRRRRRRRRRRRRRRR  
(R)(20)  
74: RGRGR RRRRRRRRRRRRRRRRRRRRRRRRR  
(R)(20)  
75: GRCRR RRRRRRRRRRRRRRRRRRRRRRRRR  
(R)(20)  
76: RGRRR RRRRRRRRRRRRRRRRRRRRRRRRR  
(R)(20)  
77: GRRRR RRRRRRRRRRRRRRRRRRRRRRRRR  
(R)(20)

1  
ABG26491 ck: 8179 len: 109 i Abg26491 Novel human diagnostic protein  
(R,K)(20)  
75: RRRGE KKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K)(20)  
76: RRGEE KKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K)(20)  
77: RGEKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K)(20)  
78: GEKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K)(20)  
79: EKKKK RKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K)(20)  
80: KKKRR KKKKKKKKKKKKKKKKKKKKKKKKK  
(K)(20)  
81: KKKRR KKKKKKKKKKKKKKKKKKKKKKKKK  
(K)(20)



1  
82: KKKKK KKKKKKKKKKKKKKKKK KGNLS  
(K){20}  
83: KKKKK KKKKKKKKKKKKKKKKK GNLVS  
(K){20}  
ABG26492 ck: 5234 len: 68 ! Abg26492 Novel human diagnostic protein #26  
(R,K){20}  
29: KKKEE RRRRRRRRRRRRRRRRRR RRRRG  
(R){20}  
30: KKEER RRRRRRRRRRRRRRRRRR RRRGK  
(R){20}  
31: KEERR RRRRRRRRRRRRRRRRRR RRGKK  
(R){20}  
32: EERRR RRRRRRRRRRRRRRRRRR RCKKD  
(R){20}  
33: ERRRR RRRRRRRRRRRRRRRRRR GKKG  
(R){20}  
ABG26493 ck: 4204 len: 80 ! Abg26493 Novel human diagnostic protein #26  
(R,K){20}  
27: EKEKE KRRRRRRRRRRRRRRRRR RRRR  
(R,K){20}  
28: KEKEK RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
29: EKEK RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
30: KEKR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
31: EKRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
32: KRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
33: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
34: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
35: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
36: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
37: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
38: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
39: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
40: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}

1  
41: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
42: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
43: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
44: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
45: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
46: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
47: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
48: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
49: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
50: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
51: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
52: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
53: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
ABG26496 ck: 8829 len: 90 ! Abg26496 Novel human diagnostic protein  
(R,K){20}  
61: EKKE KRRRRRRRRRRRRRRRRR RRRR  
(R,K){20}  
62: KKEK RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
63: KEEK RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
64: EKKR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
65: EKRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
66: KRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
67: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
68: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
69: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
70: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R){20}

ABG26497 Ch: 9186 Len: 115 | Abg26497 Novel human diagnostic protein #26

(R,K)(20)  
24: EGED RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
25: GEED RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
26: EDDR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
27: EDNR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
28: DRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
29: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
30: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
31: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
32: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
33: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
34: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
35: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
36: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
37: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
38: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
39: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
40: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
41: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
42: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
43: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
44: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
45: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
46: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)

(R)(20)  
47: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
48: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
49: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
50: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
51: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
52: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
53: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
54: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
55: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
56: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
57: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
58: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
59: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
60: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
61: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
62: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
63: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
64: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
65: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
66: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
67: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
68: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
69: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
70: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)  
71: RRRR RRRRRRRRRRRRRRRRRRR RR  
(R)(20)



1

```

ABG26501  ck: 1730  len: 182  ! Abg26501 Novel human diagnostic protein #26
(R,K)(20)
72: KEEKE KKKKKKKKKKKKKKKKKKKKK KKKKE
(K)(20)
73: EEKEK KKKKKKKKKKKKKKKKKKKKK KKKKE
(K)(20)
74: EKEKK KKKKKKKKKKKKKKKKKKKKK KKEEE
(K)(20)
75: KEKKK KKKKKKKKKKKKKKKKKKKKK KEEEE
(K)(20)
76: EKKKK KKKKKKKKKKKKKKKKKKKKK EEEEE
(R)(20)
133: GRRRS RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
134: RRRSR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
135: RRSRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
136: RSSRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
137: SRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
138: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
139: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
140: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
141: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
142: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
143: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
144: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
145: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
146: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
147: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
148: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
149: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
150: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)

```

1

```

ABG26502  ck: 3399  len: 101  ! Abg26502 Novel human diagnostic protein
(R,K)(20)
59: EEKEG RRRRRRRRRRRRRRRRRRRR KKKRR
(R,K)(20)
60: EEKEG RRRRRRRRRRRRRRRRRRRR KKKRR
(R,K)(20)
61: EEKEG RRRRRRRRRRRRRRRRRRRR KKKRR
(R,K)(20)
62: EGRRR RRRRRRRRRRRRRRRRRRRR KRRKK
(R,K)(20)
63: GRRRR RRRRRRRRRRRRRRRRRRRR RRRKK
(R,K)(20)
64: RRRRR RRRRRRRRRRRRRRRRRRRR KRRKK
(R,K)(20)
65: RRRRR RRRRRRRRRRRRRRRRRRRR KRRKK
(R,K)(20)
66: RRRRR RRRRRRRRRRRRRRRRRRRR KRRKK
(R,K)(20)
67: RRRRR RRRRRRRRRRRRRRRRRRRR KRRKK
(R,K)(20)
68: RRRRR RRRRRRRRRRRRRRRRRRRR KRRKK
(R,K)(20)
69: RRRRR RRRRRRRRRRRRRRRRRRRR KRRKK
(R,K)(20)

```

1

```

ABG26505  ck: 4704  len: 93  ! Abg26505 Novel human diagnostic protein
(R,K)(20)
60: KEEEG RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
61: EEKEG RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
62: EEKEG RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
63: EGRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
64: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
65: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
66: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
67: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
68: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
69: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)

```

1 ABG26506 ck: 3684 len: 85 i Abg26506 Novel human diagnostic protein #26

(R,K){20}

47: EKEKE KKKKKKKKKKKKKKKKKKK

(R,K){20}

48: KEKEK KKKKKKKKKKKKKKKKKKK

(R,K){20}

49: EKEER KKKKKKKKKKKKKKKKKKK

(R,K){20}

50: KEERK KKKKKKKKKKKKKKKKKKK

(R,K){20}

51: EKRRK KKKKKKKKKKKKKKKKKKK

(R,K){20}

52: KRKRK KKKKKKKKKKKKKKKKKKK

(K){20}

53: RKRKR KKKKKKKKKKKKKKKKKKK

(K){20}

54: KRKRK KKKKKKKKKKKKKKKKKKK

(K){20}

55: RKRKR KKKKKKKKKKKKKKKKKKK

(K){20}

56: KRKRK KKKKKKKKKKKKKKKKKKK

(K){20}

57: RKRKR KKKKKKKKKKKKKKKKKKK

(R,K){20}

58: KKKKK KKKKKKKKKKKKKKKKKKK

(R,K){20}

59: KKKKK KKKKKKKKKKKKKKKKKKK

(R,K){20}

60: KKKKK KKKKKKKKKKKKKKKKKKK

(R,K){20}

61: KKKKK KKKKKKKKKKKKKKKKKKK

(R,K){20}

62: KKKKK KKKKKKKKKKKKKKKKKKK

1 ABG26507 ck: 9838 len: 109 i Abg26507 Novel human diagnostic protein #26

(R,K){20}

60: EEEEE RRRRRRRRRRRRRRRRRRR

(R){20}

61: EEEEE RRRRRRRRRRRRRRRRRRR

(R){20}

62: EEEEE RRRRRRRRRRRRRRRRRRR

(R){20}

63: EEEEE RRRRRRRRRRRRRRRRRRR

(R,K){20}

64: EEEEE RRRRRRRRRRRRRRRRRRR

(R,K){20}

65: RRRRR RRRRRRRRRRRRRRRRRRR

1 ABG26508 ck: 4488 len: 121 i Abg26508 Novel human diagnostic protein

(R,K){20}

56: DDEEE RRRRRRRRRRRRRRRRRRR

(R,K){20}

57: DEEEER KKKKKKKKKKKKKKKKKKK

(R,K){20}

58: EEEER RKKKKKKKKKKKKKKKKKK

(R,K){20}

59: EEEER KKKKKKKKKKKKKKKKKKK

(R,K){20}

60: EERRR KKKKKKKKKKKKKKKKKKK

(R,K){20}

61: RRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}

62: RRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}

63: RKKKK KKKKKKKKKKKKKKKKKKK

(R,K){20}

64: KKKKK KKKKKKKKKKKKKKKKKKK

(R,K){20}

65: KKKKK KKKKKKKKKKKKKKKKKKK

(R,K){20}

66: KKKKK KKKKKKKKKKKKKKKKKKK

(R,K){20}

67: KKKKK KKKKKKKKKKKKKKKKKKK

(R,K){20}

68: KKKKK KKKKKKKKKKKKKKKKKKK

(R,K){20}

69: KKKKK KKKKKKKKKKKKKKKKKKK

1 ABG26510 ck: 3426 len: 74 i Abg26510 Novel human diagnostic protein

(R,K){20}

47: RRRRS RRRRRRRRRRRRRRRRRRR

(R){20}

48: RRRRS RRRRRRRRRRRRRRRRRRR

(R){20}

49: RRSRR RRRRRRRRRRRRRRRRRRR

(R){20}

50: RSRRR RRRRRRRRRRRRRRRRRRR

(R){20}

51: SRRRR RRRRRRRRRRRRRRRRRRR

1 ABG26513 ck: 3117 len: 265 i Abg26513 Novel human diagnostic protein

(R,K){20}

199: EEEEE KKKKKKKKKKKKKKKKKKK

(K){20}

222: KKEEE KKKKKKKKKKKKKKKKK KRKEE  
      (K,K){20}  
223: KKEEK KKKKKKKKKKKKKKKKK KRKEE  
      (K,K){20}  
224: EEEKK KKKKKKKKKKKKKKKKK KRKEE  
      (K,K){20}  
225: EKKKK KKKKKKKKKKKKKKKKK EEEEE  
      (K,K){20}

ABG26514 ck: 9106 len: 218 i Abg26514 Novel human diagnostic protein #26  
      (R,K){20}  
182: EERRG RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
183: ERRCR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
184: RRCRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
185: RGRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
186: GRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
187: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
188: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
189: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
190: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
191: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
192: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
193: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}

1  
ABG26515 ck: 7288 len: 389 i Abg26515 Novel human diagnostic protein #26  
      (R,K){20}  
259: KKEES RKKKKKKKKKKKKKKKK KKKRR  
      (R,K){20}  
260: KESRR KKKKKKKKKKKKKKKKK KKKRR  
      (K,K){20}  
261: EESRR KKKKKKKKKKKKKKKKK KKKRR  
      (K,K){20}  
262: ESRRK KKKKKKKKKKKKKKKKK KKKRR  
      (R,K){20}  
263: SRKKK KKKKKKKKKKKKKKKKK KKKRR  
      (R,K){20}  
264: RKKKK KKKKKKKKKKKKKKKKK RKKRR  
      (R,K){20}

265: KKKKK KKKKKKKKKKKKKKKKK RKKRR  
      (R,K){20}

1  
ABG26516 ck: 2295 len: 91 i Abg26516 Novel human diagnostic protein  
      (R,K){20}  
45: EEEEE RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
46: EEEEE RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
47: EEEEE RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
48: EEEEE RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
49: EEEEE RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
50: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
51: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
52: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
53: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
54: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}

1  
ABG26518 ck: 8431 len: 761 i Abg26518 Novel human diagnostic protein  
      (R,K){20}  
725: EEEEE RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
726: EEEEE RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
727: EEEEE RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
728: EEEEE RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
729: EEEEE RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
730: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
731: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
732: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
733: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
734: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}  
735: RRRRR RRRRRRRRRRRRRRRRR RRRRR  
      (R,K){20}

1

736: RRRR RRRRRRRRRRRRRRRR KKKK (R,K){20}

737: RRRR RRRRRRRRRRRRRRRR KKKK (R,K){20}

738: RRRR RRRRRRRRRRRRRRRR KKK (R,K){20}

739: RRRR RRRRRRRRRRRRRRRR KKN (R,K){20}

740: RRRR RRRRRRRRRRRRRRRR KN (R,K){20}

741: RRRR RRRRRRRRRRRRRR N (R,K){20}

ABG26520 ck: 2487 len: 99 ! Abg26520 Novel human diagnostic protein #26

35: EEEE RRRRRRRRRRRRRRRR RRRR (R,K){20}

36: EEEER RRRRRRRRRRRRRRRR RRRR (R){20}

37: EEEER RRRRRRRRRRRRRRRR RRRR (R){20}

38: EEEER RRRRRRRRRRRRRRRR RRRR (R){20}

39: EEEER RRRRRRRRRRRRRRRR RRRR (R){20}

40: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

41: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

42: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

43: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

44: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

45: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

46: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

47: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

48: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

49: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

50: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

51: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

52: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

53: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

54: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

55: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

56: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

57: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

58: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

59: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

60: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

61: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

62: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

63: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

64: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

65: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

66: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

67: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

68: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

69: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

70: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

71: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

72: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

73: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

74: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

75: RRRR RRRRRRRRRRRRRRRR RRRR (R){20}

(R)(20)  
76: RRRR RRRRRRRRRRRRRRRR RRI  
(R)(20)  
77: RRRR RRRRRRRRRRRRRRRR RKI  
(R)(20)  
78: RRRR RRRRRRRRRRRRRRRR KI  
(R,K)(20)  
79: RRRR RRRRRRRRRRRRRRRR I

ABG26521 ck: 482 len: 367 i Abg26521 Novel human diagnostic protein #26

1  
(R,K)(20)  
317: EEEG KKKKKKKKKKKKKKKK KKKR  
(K)(20)

(K)(20)  
318: EEEG KKKKKKKKKKKKKKKK KKKR

(K)(20)  
319: EEGK KKKKKKKKKKKKKKKK KRRR

(K)(20)  
320: EGKK KKKKKKKKKKKKKKKK KRRR

(K)(20)  
321: GKKK KKKKKKKKKKKKKKKK RRKK

(R,K)(20)  
322: KKKK KKKKKKKKKKKKKKKK RRKK

(R,K)(20)  
323: KKKK KKKKKKKKKKKKKKKK RRKK

(R,K)(20)  
324: KKKK KKKKKKKKKKKKKKKK KKKK

(R,K)(20)  
325: KKKK KKKKKKKKKKKKKKKK KKKK

(R,K)(20)  
326: KKKK KKKKKKKKKKKKKKKK KKKK

(R,K)(20)  
327: KKKK KKKKKKKKKKKKKKKK KKKK

(R,K)(20)  
328: KKKK KKKKKKKKKKKKKKKK KKKK

ABG26522 ck: 2060 len: 152 i Abg26522 Novel human diagnostic protein #26

1  
(R,K)(20)  
46: EEEE RRRRRRRRRRRRRRRR RRKE  
(R)(20)

(R)(20)  
47: EEEER RRRRRRRRRRRRRRRR RKEG

(R)(20)  
48: EEEER RRRRRRRRRRRRRRRR RKEG

(R,K)(20)  
49: EERR RRRRRRRRRRRRRRRR KEGE

(R,K)(20)  
50: EERR RRRRRRRRRRRRRRRR EEEG

(R)(20)  
96: KKEG RRRRRRRRRRRRRRRR RRRR

(R)(20)  
97: KEEG RRRRRRRRRRRRRRRR RRRR  
(R)(20)  
98: EEEG RRRRRRRRRRRRRRRR RRRR  
(R)(20)  
99: EERR RRRRRRRRRRRRRRRR RRRR  
(R)(20)  
100: GRRR RRRRRRRRRRRRRRRR RRRR

(R)(20)  
101: RRRR RRRRRRRRRRRRRRRR RRRR

(R)(20)  
102: RRRR RRRRRRRRRRRRRRRR RRRR

(R)(20)  
103: RRRR RRRRRRRRRRRRRRRR RRRR

(R)(20)  
104: RRRR RRRRRRRRRRRRRRRR RRRR

(R)(20)  
105: RRRR RRRRRRRRRRRRRRRR RRRR

(R)(20)  
106: RRRR RRRRRRRRRRRRRRRR RRRR

(R)(20)  
107: RRRR RRRRRRRRRRRRRRRR RRRR

(R)(20)  
108: RRRR RRRRRRRRRRRRRRRR RRRR

(R)(20)  
109: RRRR RRRRRRRRRRRRRRRR RRKE

(R)(20)  
110: RRRR RRRRRRRRRRRRRRRR RRKE

(R)(20)  
111: RRRR RRRRRRRRRRRRRRRR RKEE

(R)(20)  
112: RRRR RRRRRRRRRRRRRRRR RKEE

(R,K)(20)  
113: RRRR RRRRRRRRRRRRRRRR RKEE

ABG26525 ck: 543 len: 103 i Abg26525 Novel human diagnostic protein

1  
(R,K)(20)  
3: RD RRRRRRRRRRRRRRRR EEEE  
(R,K)(20)

ABG26526 ck: 1887 len: 115 i Abg26526 Novel human diagnostic protein

1  
(R,K)(20)  
48: EEEE RRRRRRRRRRRRRRRR RRKE  
(R)(20)

(R,K)(20)  
49: EEEER RRRRRRRRRRRRRRRR RKEE

(R,K)(20)  
50: EEEER RRRRRRRRRRRRRRRR RKEE

(R,K)(20)



```
51: EERRR RRRRRRRRRRRRRRRRKKR EKKS
ABG26527 ck: 5586 len: 122 ! Abg26527 Novel human diagnostic protein #26
1 (R,K){20}
82: EEEEE KKKKKKKKKKKKKRRRRR GRRRR
(R,K){20}
ABG26528 ck: 2237 len: 215 ! Abg26528 Novel human diagnostic protein #26
1 (R,K){20}
87: EKEKE KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
88: KEKEK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
89: EKEKK RKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
90: KEKKR KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
91: EKRRK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
92: KKRKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
93: KRKKK RKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
94: RKKRR KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
95: KKKRK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
96: KKRKK KKKKKKKKKKKKKKKKKKK EEEEE
(K){20}
167: GRRRG RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
168: RRRGR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
169: RRGRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
170: RGRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
171: GRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
172: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
173: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
174: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
175: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
176: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
```

```
177: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
ABG26530 ck: 5729 len: 404 ! Abg26530 Novel human diagnostic protein
1 (R,K){20}
366: EEEEE RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
367: EEEEE RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
368: EERRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
369: EERRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
370: ERRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
371: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
372: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
373: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
374: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
375: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
376: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
377: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
378: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
379: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
ABG26531 ck: 7434 len: 126 ! Abg26531 Novel human diagnostic protein
1 (R,K){20}
48: EEKDE RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
49: EKDER RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
50: KDERR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
51: DERRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
52: ERRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
53: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
54: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
```

```

57: RRRR RRRRRRRRRRRRRRRRRR EEEEE
56: RRRR RRRRRRRRRRRRRRRRRR REEEE
55: RRRR RRRRRRRRRRRRRRRRRR RREEE
54: RRRR RRRRRRRRRRRRRRRRRR RRRRR
53: RRRR RRRRRRRRRRRRRRRRRR RRRRR
52: RRRR RRRRRRRRRRRRRRRRRR RRRRR
51: RRRR RRRRRRRRRRRRRRRRRR RRRRR
50: RRRR RRRRRRRRRRRRRRRRRR RRRRR
49: RRRR RRRRRRRRRRRRRRRRRR RRRRR
48: RRRR RRRRRRRRRRRRRRRRRR RRRRR
47: RRRR RRRRRRRRRRRRRRRRRR RRRRR
46: RRRR RRRRRRRRRRRRRRRRRR RRRRR
45: RRRR RRRRRRRRRRRRRRRRRR RRRRR
44: RRRR RRRRRRRRRRRRRRRRRR RRRRR
43: RRRR RRRRRRRRRRRRRRRRRR RRRRR
42: RRRR RRRRRRRRRRRRRRRRRR RRRRR
41: RRRR RRRRRRRRRRRRRRRRRR RRRRR
40: RRRR RRRRRRRRRRRRRRRRRR RRRRR
39: RRRR RRRRRRRRRRRRRRRRRR RRRRR
38: RRRR RRRRRRRRRRRRRRRRRR RRRRR
37: RRRR RRRRRRRRRRRRRRRRRR RRRRR
36: RRRR RRRRRRRRRRRRRRRRRR RRRRR
35: RRRR RRRRRRRRRRRRRRRRRR RRRRR
34: RRRR RRRRRRRRRRRRRRRRRR RRRRR
33: RRRR RRRRRRRRRRRRRRRRRR RRRRR
32: RRRR RRRRRRRRRRRRRRRRRR RRRRR
31: RRRR RRRRRRRRRRRRRRRRRR RRRRR
30: RRRR RRRRRRRRRRRRRRRRRR RRRRR
29: RRRR RRRRRRRRRRRRRRRRRR RRRRR
28: RRRR RRRRRRRRRRRRRRRRRR RRRRR
27: RRRR RRRRRRRRRRRRRRRRRR RRRRR
26: RRRR RRRRRRRRRRRRRRRRRR RRRRR
25: RRRR RRRRRRRRRRRRRRRRRR RRRRR
24: RRRR RRRRRRRRRRRRRRRRRR RRRRR
23: RRRR RRRRRRRRRRRRRRRRRR RRRRR
22: RRRR RRRRRRRRRRRRRRRRRR RRRRR
21: RRRR RRRRRRRRRRRRRRRRRR RRRRR
20: RRRR RRRRRRRRRRRRRRRRRR RRRRR
19: RRRR RRRRRRRRRRRRRRRRRR RRRRR
18: RRRR RRRRRRRRRRRRRRRRRR RRRRR
17: RRRR RRRRRRRRRRRRRRRRRR RRRRR
16: RRRR RRRRRRRRRRRRRRRRRR RRRRR
15: RRRR RRRRRRRRRRRRRRRRRR RRRRR
14: RRRR RRRRRRRRRRRRRRRRRR RRRRR
13: RRRR RRRRRRRRRRRRRRRRRR RRRRR
12: RRRR RRRRRRRRRRRRRRRRRR RRRRR
11: RRRR RRRRRRRRRRRRRRRRRR RRRRR
10: RRRR RRRRRRRRRRRRRRRRRR RRRRR
9: RRRR RRRRRRRRRRRRRRRRRR RRRRR
8: RRRR RRRRRRRRRRRRRRRRRR RRRRR
7: RRRR RRRRRRRRRRRRRRRRRR RRRRR
6: RRRR RRRRRRRRRRRRRRRRRR RRRRR
5: RRRR RRRRRRRRRRRRRRRRRR RRRRR
4: RRRR RRRRRRRRRRRRRRRRRR RRRRR
3: RRRR RRRRRRRRRRRRRRRRRR RRRRR
2: RRRR RRRRRRRRRRRRRRRRRR RRRRR
1: RRRR RRRRRRRRRRRRRRRRRR RRRRR

```

1

ABG26532	ck: 9650	len: 225	! Abg26532 Novel human diagnostic protein #36
(R,K){20}			
(R,K){20}			

```

152: EEEEE KKKRRRRRRRRRRRRRRRR RRRR
      (R,K) (20)
153: EEEEE KKKRRRRRRRRRRRRRRRR RRRR

```

```
(R) { 20 }
155: EEKK RRRRRRRRRRRRRRRRR RR
```

```

157: KKKR RRRRRRRRRRRRRRRRRRR RRRR
      (R) (20)

```

```

159: KRRRR RRRRRRRRRRRRRRRRRRRRRRRRRR RRRRR
      (R) { 20 }

```

```
(R) (20)
161: 00000 000000000000000000 00000
```

(R) {20}

$$(R)_{\{20\}}$$

(B) {201

Accession	Length	Accession	Length	Accession	Length	Accession	Length
ABG26533	5900	len: 101	ABG26533	Novel	human	diagnostic	protein #26
(P K) (20)							

51: RRRRT RRRRRRRRRRRRRRRRRR RRRR  
 (R) [20]  
 52: RRRTR RRRRRRRRRRRRRRRRRR RRRR

```

53: RTRR RRRRRRRRRRRRRRRRRR RRRK
      (R) (20)
54: RTRR RRRRRRRRRRRRRRRRRR RKKK
      (R) (20)
55: TRRR RRRRRRRRRRRRRRRRRR RKKP
      (R) (20)
56: RRRR RRRRRRRRRRRRRRRRRR KKKPT
      (R) (20)
57: RRRR RRRRRRRRRRRRRRRRRR KKPTS
      (R, K) (20)
58: RRRR RRRRRRRRRRRRRRRRRR KPTSR
      (R, K) (20)
59: RRRR RRRRRRRRRRRRRRRRRR PTSSV
      (R, K) (20)

```

ABG26534 ck: 1919 len: 113 1 Abg26534 Novel human diagnostic protein  
(R K)(20)

```

(R,K) {20}
KKKKK
31: KKKKE KKKKKRRRRKKKKKKKKKKKKKK KKKKK

```

33: KEEKK RKKRRKKKKKKKKKKKKKK KKKKK  
(R,K) (20)

```
(R,K){20}
35. EKKRK KDKDKKXXXXXXXXXXXX
```

(R, K) {20}

 $(R, K) \{20\}$ 

..... (K) (20)

42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

44: RKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

47: KKKK KKKKKKKKKKKKKKK KKKK  
(K){20}  
48: KKKK KKKKKKKKKKKKKKK KKKK  
(K){20}  
49: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
50: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
51: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
52: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
53: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
54: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
55: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}

ABG26535 ck: 2214 len: 472 1 Abg26535 Novel human diagnostic protein #26  
(R,K){20}  
415: EEEE RRRRRRRRRRRRRRR RRRR  
(R){20}  
416: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
417: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
418: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
419: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
420: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
421: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
422: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
423: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
424: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
425: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
426: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
427: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
428: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}

429: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
430: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
431: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
432: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
433: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
434: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
435: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
436: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
437: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
438: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}

ABG26537 ck: 2429 len: 573 1 Abg26537 Novel human diagnostic protein  
(R,K){20}  
405: GAEE RRRRRRRRRRRRRRR RREE  
(R){20}  
406: AEDE RRRRRRRRRRRRRRR REEE  
(R){20}  
407: EDEE RRRRRRRRRRRRRRR EEEE  
(R){20}  
479: KEEE KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
480: KEEE KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
481: EEEE KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
482: EEEE KKKKKKKKKKKKKKK KKKK  
(K){20}  
483: EKKK KKKKKKKKKKKKKKK KKKK  
(K){20}  
484: KKKK KKKKKKKKKKKKKKK KKKK  
(K){20}  
485: KKKK KKKKKKKKKKKKKKK KKKK  
(K){20}  
486: KKKK KKKKKKKKKKKKKKK KKKK  
(K){20}

ABG26538 ck: 5732 len: 98 1 Abg26538 Novel human diagnostic protein  
(R,K){20}  
60: KEKE KRRRRRRRRRRRRRR KKKK  
(R,K){20}

```

61: EKETK RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
62: KETKR RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
63: ETKRR RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
64: TKRRR RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
65: KRRRR RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
66: RRRRR RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
67: RRRRK RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
68: RRRKK RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
69: RRRKK RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
70: RKKKK RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
71: KKKKK RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
72: KKKKR RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
73: KKKRK RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
74: KKRKK RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
75: KRKKK RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
76: RKKKK RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}
77: KKKKK RRRRRKKKKKKKKKKKKKKKKKK (R,K){20}

```

```

1
ABG26539 ck: 3545 len: 89 i Abg26539 Novel human diagnostic protein #26
(R,K){20}
(R,K){20}
28: EEEEE RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
29: EEEER RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
30: EEEER RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
31: EEEER RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
32: EEEER RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
33: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}

```

```

34: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
35: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
36: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
37: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
38: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
39: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
40: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
41: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
42: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
43: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
44: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
45: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
46: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
47: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
48: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
49: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
50: RRRRR RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}

```

```

1
ABG26542 ck: 4605 len: 182 i Abg26542 Novel human diagnostic protein
(R,K){20}
(R,K){20}
146: EEEEE RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
147: GEEER RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
148: EEEER RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
149: EEEER RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}
ABG26543 ck: 2998 len: 405 i Abg26543 Novel human diagnostic protein
(R,K){20}
(R,K){20}
264: RRRRW RRRRRRRRRRRRRRRRRRRRRRR (R,K){20}

```



18: KKKKK (R,K){20}  
19: KKKKK (R,K){20}  
20: KKKKK (R,K){20}  
21: KKKKK (R,K){20}

ABG26717 ck: 4415 len: 78 ! Abg26717 Novel human diagnostic protein #26

37: QMLSV (R,K){20}  
38: MLSVK (R,K){20}

39: LSVKK (R,K){20}  
40: SVKKK (R,K){20}

41: VKKKK (R,K){20}

42: KKKKK (R,K){20}

43: KKKKK (R,K){20}

44: KKKKK (R,K){20}

45: KKKKK (R,K){20}

46: KKKKK (R,K){20}

47: RKKKK (R,K){20}

48: KKKKK (R,K){20}

49: KKKKK (R,K){20}

50: KKKKK (R,K){20}

51: KKKKK (R,K){20}

52: KKKKK (R,K){20}

53: KKKKK (R,K){20}

54: KKKKK (R,K){20}

ABG26718 ck: 9531 len: 141 ! Abg26718 Novel human diagnostic protein #26

1

80: EVARP (R,K){20}

81: VARP (R,K){20}

82: ARPR (R,K){20}

83: RPRKK (R,K){20}

84: PRKKK (R,K){20}

85: RKKKK (R,K){20}

86: KKKKK (R,K){20}

87: KKKKK (R,K){20}

88: KKKKK (R,K){20}

89: KKKKK (R,K){20}

ABG26719 ck: 72 len: 83 ! Abg26719 Novel human diagnostic protein

42: ETPSE (R,K){20}

43: TPSEK (R,K){20}

44: PSEKK (R,K){20}

45: SEKKK (R,K){20}

46: EKKKK (R,K){20}

47: KKKKK (R,K){20}

48: KKKKK (R,K){20}

49: KKKKK (R,K){20}

50: KKKKK (R,K){20}

51: KKKKK (R,K){20}

52: KKKKK (R,K){20}

53: KKKKK (R,K){20}

(K){20}

1

54: KKKKK KKKKKKKKKKKKKKKKKKKKKK NPIEF  
ABG26720 ck: 3781 len: 57 ! Abg26720 Novel human diagnostic protein #26  
(R,K){20}  
24: QMKSI KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: MKSIK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: KSIKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
27: SIKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
28: IKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
29: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
31: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
32: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
33: KKKKK KKKKKKKKKKKKKKKKKKKKKK FHEFL  
(K){20}

1

ABG26721 ck: 287 len: 95 ! Abg26721 Novel human diagnostic protein #26  
(R,K){20}  
31: EKEKE KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
32: KEKEK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
33: EKEKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
34: KEKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
35: EKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
36: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
37: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
38: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
39: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
40: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
41: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

1

42: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
43: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
44: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
45: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
46: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
50: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
51: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKKKKKKK NKKKK  
(K){20}  
ABG26722 ck: 4831 len: 127 ! Abg26722 Novel human diagnostic protein  
(R,K){20}  
44: KKKEE KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
45: KKEEK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
46: KEEKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
47: EEKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: EKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
87: EKEEE KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
88: KEEEK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
89: EEEKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
90: EEKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

91: EKKKK (K){20} KKKKK  
92: KKKKK (K){20} KKKKK  
93: KKKKK (K){20} KKKKK  
94: KKKKK (K){20} KKKKK  
95: KKKKK (K){20} KKKKK  
96: KKKKK (K){20} KKKKK  
97: KKKKK (K){20} KKKKK  
98: KKKKK (K){20} KKKKK  
99: KKKKK (K){20} KKKKK  
100: KKKKK (K){20} KKKKK  
101: KKKKK (K){20} KKKKK  
102: KKKKK (K){20} KKKKK  
103: KKKKK (K){20} KKKKK  
104: KKKKK (K){20} KKKKK  
105: KKKKK (K){20} KKKKK  
106: KKKKK (K){20} KKKKK  
107: KKKKK (K){20} KKKKK  
108: KKKKK (K){20} KKKKK  
ABG26723 ck: 7054 len: 98 1 Abg26723 Novel human diagnostic protein #26  
(R,K){20}  
42: EKEKE (K){20} KKKKK  
43: KEKEK (K){20} KKKKK  
44: EKEKK (K){20} KKKKK  
45: KEKKK (K){20} KKKKK  
46: EKKKK (K){20} KKKKK

47: KKKKK (K){20} KKKKK  
48: KKKKK (K){20} KKKKK  
49: KKKKK (K){20} KKKKK  
50: KKKKK (K){20} KKKKK  
51: KKKKK (K){20} KKKKK  
52: KKKKK (K){20} KKKKK  
53: KKKKK (K){20} KKKKK  
54: KKKKK (K){20} KKKKK  
55: KKKKK (K){20} KKKKK  
56: KKKKK (K){20} KKKKK  
57: KKKKK (K){20} KKKKK  
58: KKKKK (K){20} KKKKK  
59: KKKKK (K){20} KKKKK  
60: KKKKK (K){20} KKKKK  
61: KKKKK (K){20} KKKKK  
62: KKKKK (K){20} KKKKK  
63: KKKKK (K){20} KKKKK  
64: KKKKK (K){20} KKKKK  
65: KKKKK (K){20} KKKKK  
66: KKKKK (K){20} KKKKK  
67: KKKKK (K){20} KKKKK  
68: KKKKK (K){20} KKKKK  
69: KKKKK (K){20} KKKKK  
70: KKKKK (K){20} KKKKK



1 ABG26724 ck: 9897 len: 295 ! Abg26724 Novel human diagnostic protein #26

(R,K){20}

259: KEEEE KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

260: EEEER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

261: EEEER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

262: EEEER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

263: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

264: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

265: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

266: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

267: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

268: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

269: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

270: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

271: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

272: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

273: KKKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

274: KKKKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

275: KKKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

276: KKKKK KKKKKKKKKKKKKKKKKKK

1 ABG26725 ck: 909 len: 131 ! Abg26725 Novel human diagnostic protein #26

(R,K){20}

70: KKKKE KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

71: KKKKE KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

72: KKKKE KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

73: KKKR KKKKKKKKKKKKKKKKKKK KKKKK

1

74: EKKKK (K){20}

KKKKKKKKKKKKKKKKKKKK KKKKK

75: KKKKK (K){20}

KKKKKKKKKKKKKKKKKKKK KKKKK

76: KKKKK (K){20}

KKKKKKKKKKKKKKKKKKKK KKKKK

77: KKKKK (K){20}

KKKKKKKKKKKKKKKKKKKK KKKKK

78: KKKKK (K){20}

KKKKKKKKKKKKKKKKKKKK KKKKK

79: KKKKK (K){20}

KKKKKKKKKKKKKKKKKKKK KKKKK

80: KKKKK (K){20}

KKKKKKKKKKKKKKKKKKKK KKKKK

81: KKKKK (K){20}

KKKKKKKKKKKKKKKKKKKK KKKKK

82: KKKKK (K){20}

KKKKKKKKKKKKKKKKKKKK KKKKK

83: KKKKK (K){20}

KKKKKKKKKKKKKKKKKKKK KKKKK

ABG26726 ck: 9300 len: 677 ! Abg26726 Novel human diagnostic protein

(R,K){20}

240: EKEKE (R,K){20}

KKKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

241: KEKER KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

242: EKEER KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

243: KEKER KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

244: EKKER KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

245: KKKER KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

246: KKKER KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

247: KKKER KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

248: KKKER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

249: KKKER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

250: KKKER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

251: KKKER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

252: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

1

253: RKKKK (K){20} KKKKK  
254: KKKKK (K){20} KKKKK  
255: KKKKK (K){20} KKKKK  
256: KKKKK (K){20} KKKKK  
257: KKKKK (K){20} KKKKK  
258: KKKKK (K){20} KKKKK  
259: KKKKK (K){20} KKKKK  
260: KKKKK (K){20} KKKKK  
261: KKKKK (R,K){20} KKKKK  
262: KKKKK (R,K){20} KKKKK  
263: KKKKK (R,K){20} KKKKK  
264: KKKKK (R,K){20} KKKKK  
265: KKKKK (R,K){20} KKKKK  
ABG26727 ck: 813 len: 329 1 Abg26727 Novel human diagnostic protein #26

(R,K){20}  
(K){20}  
(K){20}  
(K){20}  
(K){20}  
(K){20}  
(K){20}  
(K){20}  
(K){20}  
(K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}

1

225: KKKKK (K){20} KKKKK  
226: KKKKK (K){20} KKKKK  
227: KKKKK (K){20} KKKKK  
228: KKKKK (K){20} KKKKK  
229: KKKKK (K){20} KKKKK  
230: KKKKK (K){20} KKKKK  
ABG26879 ck: 2669 len: 800 1 Abg26879 Novel human diagnostic protein

(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}

1

AAG65985 ck: 8085 len: 154 1 Aag65985 B726P splice variant sequence.

(R,K){20}  
(K){20}

115: QLRQK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
116: LROKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
117: ROQKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
118: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
119: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
120: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
121: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
122: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
123: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
124: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
125: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
126: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
127: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
128: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

1  
ABb27893 ck: 5383 len: 86 ! Abb27893 Human peptide #544 encoded by brea  
(R,K){20}  
15: RRRRG RRRRRKKKKKKKKKKKKKK KKKRR  
(R,K){20}  
16: RRRGR RRRRRKKKKKKKKKKKKKK KKKRR  
(R,K){20}  
17: RRGRR RRRRRKKKKKKKKKKKKKK KKKRR  
(R,K){20}  
18: RGRRR RRRRRKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
19: GRRRR RRRRRKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
20: RRRRR KKKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
21: RRRRK KKKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
22: RRRKK KKKRRKKKKKKKKKKRRRR RRRRR  
(R,K){20}  
23: RRRKK KKKKKKKKKKKKKRRRRRR RRRRR  
(R,K){20}

1  
24: RKKKK KRRKKKKKKKKRRRRRRRR RRRRR  
(R,K){20}  
25: KKKKK KRRKKKKKKKKRRRRRRRR RRRRR  
(R,K){20}  
26: KKKKK RKKKKKKKKKKRRRRRRRR RRRRR  
(R,K){20}  
27: KKKKK KKKKKKKKKKKRRRRRRRR RRRRR  
(R,K){20}  
28: KKKKK KKKKKKKKKKKRRRRRRRR RRRRR  
(R,K){20}  
29: KKKKK KKKKKKKKKKKRRRRRRRR RRRRR  
(R,K){20}  
30: KRRKK KRRKKKKKKKKRRRRRRRR RRRRR  
(R,K){20}  
31: RKKKK KRRKKKKKKKKRRRRRRRR RRRRR  
(R,K){20}  
32: KKKKK RKKKKKKKKKKRRRRRRRR RRRRR  
(R,K){20}  
33: KKKKK KRRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
34: KKKKK KRRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
35: KRRKK KRRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
36: KRRKK KRRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
37: RKKKK RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
38: KKKKK RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
39: KKKKK RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
40: KRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
41: KRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
42: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
43: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
44: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
45: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}

ABb28750 ck: 1334 len: 86 ! Abb28750 Peptide #1401 encoded by brea  
(R,K){20}  
57: EEEEG RRRKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}



52: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
53: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
54: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
55: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
56: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
57: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
58: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
59: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
60: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
61: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
62: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
63: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
64: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
65: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
66: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
67: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
68: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
69: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK

ABB30512 ck: 3937 len: 85 ! Abb30512 Peptide #3163 encoded by breast ce

1: (R,K){20} KKKKKKKKKKKKKKKKKKK KKKK  
2: (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
3: (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
4: (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
5: (K){20} KKKKKKKKKKKKKKKKKKK KKKK

6: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
7: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
8: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
9: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
10: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK  
11: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK

1

ABB32308 ck: 1560 len: 88 ! Abb32308 Peptide #4959 encoded by breas

43: EERG (R,K){20} RRRRRRRRRRRRRRRRRR RRRR  
44: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
45: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
46: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
47: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
48: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
49: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
50: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
51: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
52: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
53: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
54: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
55: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
56: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
57: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
58: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR  
59: EERG (R){20} RRRRRRRRRRRRRRRRRR RRRR

60: RRRR (R,20) RRRRRRRRRRRRRRRRRR RRNTN  
61: RRRR (R,20) RRRRRRRRRRRRRRRRRR RNTNN  
62: RRRR (R,20) RRRRRRRRRRRRRRRRRR NTNNE  
ABB31064 ck: 5383 len: 86 1 Abb31064 Peptide #570 encoded by human foet

15: RRRG (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
16: RRGR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
17: RGRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
18: RGRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
19: GRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
20: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
21: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
22: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
23: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
24: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
25: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
26: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
27: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
28: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
29: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
30: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
31: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
32: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
33: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR  
34: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR KRRR

35: KRRR (R,K){20} KRRRRRRRRRRRRRRRRRRRRRRRRR  
36: KRRR (R,K){20} KRRRRRRRRRRRRRRRRRRRRRRRRR  
37: KRRR (R,20) KRRRRRRRRRRRRRRRRRRRRRRRRR  
38: KRRR (R,20) KRRRRRRRRRRRRRRRRRRRRRRRRR  
39: KRRR (R,20) KRRRRRRRRRRRRRRRRRRRRRRRRR  
40: KRRR (R,20) KRRRRRRRRRRRRRRRRRRRRRRRRR  
41: KRRR (R,20) KRRRRRRRRRRRRRRRRRRRRRRRRR  
42: KRRR (R,20) KRRRRRRRRRRRRRRRRRRRRRRRRR  
43: KRRR (R,20) KRRRRRRRRRRRRRRRRRRRRRRRRR  
44: KRRR (R,20) KRRRRRRRRRRRRRRRRRRRRRRRRR  
45: KRRR (R,20) KRRRRRRRRRRRRRRRRRRRRRRRRR  
ABB3937 ck: 1334 len: 86 1 Abb3937 Peptide #1443 encoded by human

57: EEEG (R,K){20} EEEGRRRRRRRRRRRRRRRRRRRRRRRRR  
58: EEEG (R,K){20} EEEGRRRRRRRRRRRRRRRRRRRRRRRRR  
59: EEEG (R,K){20} EEEGRRRRRRRRRRRRRRRRRRRRRRRRR  
60: EEEG (R,K){20} EEEGRRRRRRRRRRRRRRRRRRRRRRRRR  
61: EEEG (R,K){20} EEEGRRRRRRRRRRRRRRRRRRRRRRRRR  
62: EEEG (R,K){20} EEEGRRRRRRRRRRRRRRRRRRRRRRRRR  
63: EEEG (R,K){20} EEEGRRRRRRRRRRRRRRRRRRRRRRRRR  
64: EEEG (R,K){20} EEEGRRRRRRRRRRRRRRRRRRRRRRRRR  
65: EEEG (R,K){20} EEEGRRRRRRRRRRRRRRRRRRRRRRRRR  
66: EEEG (R,K){20} EEEGRRRRRRRRRRRRRRRRRRRRRRRRR  
67: EEEG (R,K){20} EEEGRRRRRRRRRRRRRRRRRRRRRRRRR  
ABB34024 ck: 9082 len: 167 1 Abb34024 Peptide #1530 encoded by human

```
1
33: EECGRG RRRRRRRRRRRRRRRRRRRR RRGCG
      (R,K){20}
34: EGRGR RRRRRRRRRRRRRRRRRRRR RGGGR
      (R,K){20}
35: GRCGR RRRRRRRRRRRRRRRRRRRR GCGRR
      (R,K){20}
ABB3453 ck: 2276 len: 89 ! Abb3453 Peptide #2039 encoded by human fo
      (R,K){20}
23: EEEEE KKKKKKKKKKKKKKKKKKKK EEEKK
      (R,K){20}
      (K){20}
46: KKEEE KKKKKKKKKKKKKKKKKKKK KRKEE
      (K){20}
47: KEEBK KKKKKKKKKKKKKKKKKKKK RKEEE
      (K){20}
48: EEEKK KKKKKKKKKKKKKKKKKKKK KEEBE
      (R,K){20}
49: EKKKK KKKKKKKKKKKKKKKKKKKK EEEEE
      (R,K){20}
ABB34819 ck: 1939 len: 130 ! Abb34819 Peptide #2325 encoded by human fo
      (R,K){20}
42: EGKKE RRRRRRRRRRRRRRRRRRRR RRRRR
      (R,K){20}
43: GKKEE RRRRRRRRRRRRRRRRRRRR RRRRR
      (R,K){20}
44: RKEER RRRRRRRRRRRRRRRRRRRR RRRRR
      (R,K){20}
45: KEERR RRRRRRRRRRRRRRRRRRRR RRRKK
      (R,K){20}
46: ERRRR RRRRRRRRRRRRRRRRRRRR RRRKK
      (R,K){20}
47: RRRRR RRRRRRRRRRRRRRRRRRRR RKKKK
      (R,K){20}
48: RRRRR RRRRRRRRRRRRRRRRRRRR RKKKK
      (R,K){20}
49: RRRRR RRRRRRRRRRRRRRRRRRRR KKKKK
      (R,K){20}
50: RRRRR RRRRRRRRRRRRRRRRRRRR KKKKE
      (R,K){20}
51: RRRRR RRRRRRRRRRRRRRRRRRRR KKKEE
      (R,K){20}
52: RRRRR RRRRRRRRRRRRRRRRRRRR KKEEE
      (R,K){20}
53: RRRRR RRRRRRRRRRRRRRRRRRRR KEEBE
      (R,K){20}
54: RRRRR RRRRRRRRRRRRRRRRRRRR EEEEE
      (R,K){20}
```

```
1
ABB35110 ck: 3607 len: 88 ! Abb35110 Peptide #2616 encoded by human fo
      (R,K){20}
39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
40: RKREK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
41: KREKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
42: REKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
```

62: KKKKK (K){20} KKKKK  
63: KKKKK (K){20} KKKKK  
64: KKKKK (K){20} KKKKK  
65: KKKKK (K){20} KKKKK  
66: KKKKK (K){20} KKKKK  
67: KKKKK (K){20} KKKKK  
68: KKKKK (K){20} KKKKK  
69: KKKKK (K){20} KKKKK

1 ABB35676 ck: 3937 len: 85 ! Abb35676 Peptide #3182 encoded by human fog

1: (R,K){20} KKKKK  
2: (K){20} KKKKK  
3: (K){20} KKKKK  
4: (K){20} KKKKK  
5: (K){20} KKKKK  
6: (K){20} KKKKK  
7: (K){20} KKKKK  
8: (K){20} KKKKK  
9: (K){20} KKKKK  
10: (K){20} KKKKK  
11: (K){20} KKKKK

1 ABB36406 ck: 2686 len: 71 ! Abb36406 Peptide #3912 encoded by human fog

20: (R,K){20} KKKKK  
21: (K){20} KKKKK  
22: (K){20} KKKKK

23: (K){20} KKKKK  
24: (K){20} KKKKK  
25: (K){20} KKKKK  
26: (R,K){20} KKKKK  
27: (R,K){20} KKKKK  
28: (R,K){20} KKKKK  
29: (R,K){20} KKKKK  
30: (R,K){20} KKKKK  
31: (R,K){20} KKKKK  
32: (R,K){20} KKKKK  
33: (R,K){20} KKKKK  
34: (R,K){20} KKKKK  
35: (R,K){20} KKKKK  
36: (R,K){20} KKKKK  
37: (R,K){20} KKKKK  
38: (R,K){20} KKKKK  
39: (R,K){20} KKKKK  
40: (R,K){20} KKKKK  
41: (R,K){20} KKKKK  
42: (R,K){20} KKKKK  
43: (R,K){20} KKKKK  
44: (R,K){20} KKKKK  
45: (R,K){20} KKKKK  
46: (R,K){20} KKKKK



47: KKKR (K){20} KKKKKKKKKKKKKKK KSAH  
48: KKKR (K){20} KKKKKKKKKKKKKKK KSAH  
49: KKKR (K){20} KKKKKKKKKKKKKKK SAH

ABb37567 ck: 1560 len: 88 ! Abb37567 Peptide #5073 encoded by human fo

43: RRRG (R,K){20} RRRRRRRRRRRRRRRR RRRR  
44: ERGR (R){20} RRRRRRRRRRRRRRRR RRRR  
45: RGGR (R){20} RRRRRRRRRRRRRRRR RRRR  
46: RGRR (R){20} RRRRRRRRRRRRRRRR RRRR  
47: GRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
48: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
49: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
50: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
51: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
52: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
53: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
54: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
55: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
56: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
57: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
58: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
59: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
60: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
61: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
62: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR

ABb37780 ck: 2324 len: 36 ! Abb37780 Peptide #5286 encoded by huma  
16: KERK (R,K){20} KKKRRRRRRRRRRRRRR R  
17: ERKT (R,K){20} KKKRRRRRRRRRRRRRR

ABb40272 ck: 8343 len: 66 ! Abb40272 Peptide #7778 encoded by huma  
6: ETER (R,K){20} KKKRRRRRRRRRRRRRR KKKK  
7: TERK (R,K){20} KKKRRRRRRRRRRRRRR KKKK  
8: EREK (R,K){20} KKKRRRRRRRRRRRRRR KKKK  
9: REKK (R,K){20} KKKRRRRRRRRRRRRRR KKKK

ABb42642 ck: 2394 len: 57 ! Abb42642 Peptide #10148 encoded by huma

20: EEEG (R,K){20} RRRRRRRRRRRRRRRR RRRR  
21: EEGR (R){20} RRRRRRRRRRRRRRRR RRRR  
22: EEEG (R){20} RRRRRRRRRRRRRRRR RRRR  
23: EGRR (R){20} RRRRRRRRRRRRRRRR RRRR  
24: GRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
25: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR

ABb43181 ck: 4228 len: 24 ! Abb43181 Peptide #10687 encoded by huma

1: RRRR (R,K){20} RRRRRRRRRRRRRRRR RRRR  
2: RRRR (R,K){20} RRRRRRRRRRRRRRRR RRRR  
3: RR (R,K){20} RRRRRRRRRRRRRRRR RRRR  
4: RRR (R,K){20} RRRRRRRRRRRRRRRR RRRR

ABb44317 ck: 4895 len: 51 ! Abb44317 Peptide #11823 encoded by huma

18: LKPM (R,K){20} RRRRRRRRRRRRRRRR RRRR  
19: FKPM (R,K){20} RRRRRRRRRRRRRRRR RRRR

20: KPMRK RRRKKRRRRKKKKRRRK LTTT  
(R,K){20}

ABb17165 ck: 8887 len: 42 i Abb17165 Human nervous system related polyf

1  
(R,K){20}

21: FTTTE KKKKKKKKKKKKKKKKK KX  
(K){20}

22: FTTEK KKKKKKKKKKKKKKKKK X  
(K){20}

ABb18534 ck: 5383 len: 86 i Abb18534 Protein #533 encoded by probe for

1  
(R,K){20}

15: RRRRG RRRKKKKKKKKKKKKKK KRRR  
(R,K){20}

16: RRRGR RRRKKKKKKKKKKKKKK KRRR  
(R,K){20}

17: RRGRR RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

18: RGRRR RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

19: GRRRR RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

20: RRRRR RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

21: RRRRK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

22: RRRRK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

23: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

24: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

25: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

26: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

27: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

28: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

29: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

30: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

31: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

32: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

(R,K){20}

33: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

34: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

35: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

36: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

37: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R){20}

38: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R){20}

39: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R){20}

40: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R){20}

41: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R){20}

42: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R){20}

43: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R){20}

44: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R){20}

45: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R){20}

ABb19373 ck: 1334 len: 86 i Abb19373 Protein #1372 encoded by probe

1

(R,K){20}

57: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

58: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

59: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

60: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

61: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

62: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

63: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

64: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

65: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

66: RRRKK RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

(R,K){20}

67: KKKR KKKKKKKKKKKKKKK

ABBI9467 ck: 9082 len: 167 ! Abbi9467 Protein #1466 encoded by probe for

(R,K){20}

33: EEEG RRRRRRRRRRRRRRRR RRGCG

(R,K){20}

34: EGGR RRRRRRRRRRRRRRRR RGGGR

(R,K){20}

35: GGGRR RRRRRRRRRRRRRRRR GGGRR

ABBI9943 ck: 2276 len: 89 ! Abbi9943 Protein #1942 encoded by probe for

(R,K){20}

23: EEEEE KKKKKKKKKKKKKKKK EEEKK

(K){20}

46: KKEE KKKKKKKKKKKKKKKK KKEE

(K){20}

47: KEEK KKKKKKKKKKKKKKKK KKEE

(R,K){20}

48: EEKK KKKKKKKKKKKKKKKK KEEE

(R,K){20}

49: EEKK KKKKKKKKKKKKKKKK EEEEE

ABBI20238 ck: 1939 len: 130 ! Abbi20238 Protein #2237 encoded by probe for

(R,K){20}

42: EGKE RRRRRRRRRRRRRRRR RRRRR

(R,K){20}

43: GKRR RRRRRRRRRRRRRRRR RRRRR

(R,K){20}

44: KEER RRRRRRRRRRRRRRRR RRRRR

(R,K){20}

45: KEER RRRRRRRRRRRRRRRR RRRRR

(R,K){20}

46: ERER RRRRRRRRRRRRRRRR RRRRR

(R,K){20}

47: RRRR RRRRRRRRRRRRRRRR RRRRR

(R,K){20}

48: RRRR RRRRRRRRRRRRRRRR RRRRR

(R,K){20}

49: RRRR RRRRRRRRRRRRRRRR RRRRR

(R,K){20}

50: RRRR RRRRRRRRRRRRRRRR RRRRR

(R,K){20}

51: RRRR RRRRRRRRRRRRRRRR RRRRR

(R,K){20}

52: RRRR RRRRRRRRRRRRRRRR RRRRR

(R,K){20}

53: RRRR RRRRRRRRRRRRRRRR RRRRR

1

(R,K){20}

54: RRRR RRRRRRRRRRRRRRRR EEEEE

ABBI20531 ck: 3607 len: 88 ! Abbi20531 Protein #2530 encoded by probe

(R,K){20}

39: ERRE KKKKKKKKKKKKKKKK KKKK

(K){20}

40: RREK KKKKKKKKKKKKKKKK KKKK

(K){20}

41: RREK KKKKKKKKKKKKKKKK KKKK

(K){20}

42: REKK KKKKKKKKKKKKKKKK KKKK

(K){20}

43: EKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

44: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

45: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

46: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

47: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

48: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

49: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

50: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

51: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

52: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

53: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

54: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

55: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

56: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

57: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

58: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

59: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}

60: KKKK KKKKKKKKKKKKKKKK KKKK

61: KKKKK (K){20} KKKKK  
62: KKKKK (K){20} KKKKK  
63: KKKKK (K){20} KKKKK  
64: KKKKK (K){20} KKKKK  
65: KKKKK (K){20} KKKKK  
66: KKKKK (K){20} KKKKK  
67: KKKKK (K){20} KKKKK  
68: KKKKK (K){20} KKKKK  
69: KKKKK (K){20} KKKKK

ABb21105 ck: 3937 len: 85 i Abb21105 Protein #3104 encoded by probe for

1

1: KKKKK (R,K){20} KKKKK  
2: KKKKK (K){20} KKKKK  
3: KKKKK (K){20} KKKKK  
4: KKKKK (K){20} KKKKK  
5: KKKKK (K){20} KKKKK  
6: KKKKK (K){20} KKKKK  
7: KKKKK (K){20} KKKKK  
8: KKKKK (K){20} KKKKK  
9: KKKKK (K){20} KKKKK  
10: KKKKK (K){20} KKKKK  
11: KKKKK (K){20} KKKKK

ABb21763 ck: 2686 len: 71 i Abb21763 Protein #3762 encoded by probe for

1

20: KKKKK (R,K){20} KKKKK  
KKKKK (K){20} KKKKK

(K){20}

21: KKKKK KKKKK  
22: KKKKK (K){20} KKKKK  
23: KKKKK (K){20} KKKKK  
24: KKKKK (K){20} KKKKK  
25: KKKKK (R,K){20} KKKKK  
26: KKKKK (R,K){20} KKKKK  
27: KKKKK (R,K){20} KKKKK  
28: KKKKK (R,K){20} KKKKK  
29: KKKKK (R,K){20} KKKKK  
30: KKKKK (R,K){20} KKKKK  
31: KKKKK (R,K){20} KKKKK  
32: KKKKK (R,K){20} KKKKK  
33: KKKKK (R,K){20} KKKKK  
34: KKKKK (R,K){20} KKKKK  
35: KKKKK (R,K){20} KKKKK  
36: KKKKK (R,K){20} KKKKK  
37: KKKKK (R,K){20} KKKKK  
38: KKKKK (R,K){20} KKKKK  
39: KKKKK (R,K){20} KKKKK  
40: KKKKK (R,K){20} KKKKK  
41: KKKKK (R,K){20} KKKKK  
42: KKKKK (R,K){20} KKKKK  
43: KKKKK (R,K){20} KKKKK  
44: KKKKK (R,K){20} KKKKK  
45: KKKKK (R,K){20} KKKKK

```

46: KKKKK (K,K) (20) KKKSA
47: KKKRR (K) (20) KKSAA
48: KKKKK (K) (20) KSAH
49: KKKKK (K) (20) SAH
ABB22862 ck: 1560 len: 88 ! Abd22862 Protein #4861 encoded by probe for
(R,K) (20)
43: RRRGG (R) (20) RRRRR
44: RRRGR (R) (20) RRRRR
45: RRRGR (R) (20) RRRRR
46: RRRRR (R) (20) RRRRR
47: GRRRR (R) (20) RRRRR
48: RRRRR (R) (20) RRRRR
49: RRRRR (R) (20) RRRRR
50: RRRRR (R) (20) RRRRR
51: RRRRR (R) (20) RRRRR
52: RRRRR (R) (20) RRRRR
53: RRRRR (R) (20) RRRRR
54: RRRRR (R) (20) RRRRR
55: RRRRR (R) (20) RRRRR
56: RRRRR (R) (20) RRRRR
57: RRRRR (R) (20) RRRRR
58: RRRRR (R) (20) RRRRR
59: RRRRR (R) (20) RRRRR
60: RRRRR (R) (20) RRRRR
61: RRRRR (R) (20) RRRRR

```

1  
 62: RRRR RRRRRRRRRRRRRRRR NTNNE  
 1  
 ABB23064 ck: 2324 len: 36 | Abb23064 Protein #5063 encoded by proba  
 16: KERKT  
 (R,K){20}  
 (R,K){20}  
 KKKRRRRRRRRRRRRRRR R  
 17: ERKTK  
 (R,K){20}  
 KKKRRRRRRRRRRRRRRR  
 1  
 ABB24685 ck: 8343 len: 66 | Abb24685 Protein #6684 encoded by proba  
 (R,K){20}  
 (R,K){20}  
 6: ETERE  
 KKKRRRRRRRRRRRRRRR KKKK  
 (R,K){20}  
 7: TEREK  
 KKKRRRRRRRRRRRRRRR KKKK  
 (R,K){20}  
 8: EREKK  
 KKKRRRRRRRRRRRRRRR KKKK  
 (R,K){20}  
 9: REKKK  
 KKKRRRRRRRRRRRRRRR KKKK  
 1  
 ABB25988 ck: 2394 len: 57 | Abb25988 Protein #7987 encoded by proba  
 (R,K){20}  
 20: EEEEG  
 RRRRRRRRRRRRRRRRRR RRRR  
 (R,K){20}  
 21: EEGR  
 RRRRRRRRRRRRRRRRRR RRRG  
 (R,K){20}  
 22: EGR  
 RRRRRRRRRRRRRRRRRR RRRG  
 (R,K){20}  
 23: EGRR  
 RRRRRRRRRRRRRRRRRR RRRG  
 (R,K){20}  
 24: GRRR  
 RRRRRRRRRRRRRRRRRR RRRG  
 (R,K){20}  
 25: RRRR  
 RRRRRRRRRRRRRRRRRR RRRR  
 1  
 ABB27176 ck: 4895 len: 51 | Abb27176 Protein #9175 encoded by proba  
 (R,K){20}  
 (R,K){20}  
 18: LKPP  
 RRRRRRRRRRRRRRRRRR KLLT  
 (R,K){20}  
 19: FKPP  
 RRRRRRRRRRRRRRRRRR KLLT  
 (R,K){20}  
 20: KPPR  
 RRRRRRRRRRRRRRRRRR LTTT  
 1  
 ABB10296 ck: 3983 len: 292 | Abb10296 Human cDNA SEQ ID NO: 604. 1/  
 (R,K){20}  
 (R,K){20}  
 273: OVFP  
 RRRRRRRRRRRRRRRRRR

1 ABB10485 ck: 7611 len: 315 i ABB10485 Human cDNA SEQ ID NO: 793. 1/2002  
(R,K){20}  
273: QVFAP RRRRRRRRRRRRRRRRRRRR KGRS  
(R,K){20}  
274: VFAPR KRRRRRRRRRRRRRRRRRRR GGRSR  
(K){20}

1 AAU21948 ck: 444 len: 66 i AAU21948 Human cardiovascular system antigen  
(R,K){20}  
35: SMFES KRRRRRRRRRRRRRRRRRRR KXGRK  
(K){20}  
36: MTFES KRRRRRRRRRRRRRRRRRRR XGRKK  
(K){20}

1 AAU22148 ck: 4665 len: 34 i AAU22148 Human cardiovascular system antigen  
(R,K){20}  
10: PELLK KRRRRRRRRRRRRRRRRRRR KKKKK  
(K){20}  
11: ELLK KRRRRRRRRRRRRRRRRRRR KKKK  
(K){20}  
12: LLLK KRRRRRRRRRRRRRRRRRRR KKK  
(K){20}  
13: LLKK KRRRRRRRRRRRRRRRRRRR KK  
(K){20}  
14: LKKK KRRRRRRRRRRRRRRRRRRR K  
(K){20}  
15: KKKK KRRRRRRRRRRRRRRRRRRR

1 AAU22186 ck: 269 len: 76 i AAU22186 Human cardiovascular system antigen  
(R,K){20}  
37: TPSRA KRRRRRRRRRRRRRRRRRRR KKKKK  
(K){20}  
38: PSRAK KRRRRRRRRRRRRRRRRRRR KKKKK  
(K){20}  
39: SRAKK KRRRRRRRRRRRRRRRRRRR KKKKK  
(K){20}  
40: RAKKK KRRRRRRRRRRRRRRRRRRR KKKKK  
(K){20}  
41: AKKKK KRRRRRRRRRRRRRRRRRRR KKKKK  
(K){20}  
42: KKKKK KRRRRRRRRRRRRRRRRRRR KKKKK  
(K){20}  
43: KKKKK KRRRRRRRRRRRRRRRRRRR KKKKI  
(K){20}  
44: KKKKK KRRRRRRRRRRRRRRRRRRR KKKIK  
(K){20}

1 AAU22374 ck: 8278 len: 53 i AAU22374 Human cardiovascular system antigen  
(R,K){20}

1 30: NCGL KRRRRRRRRRRRRRRRRRRR KKKK  
(K){20}

1 31: CGILK KRRRRRRRRRRRRRRRRRRR KKK  
(K){20}

1 32: GILKK KRRRRRRRRRRRRRRRRRRR KK  
(K){20}

1 33: ILKKK KRRRRRRRRRRRRRRRRRRR K  
(K){20}

1 34: LKKKK KRRRRRRRRRRRRRRRRRRR  
(K){20}

1 AAU27399 ck: 6158 len: 272 i AAU27399 Novel human enzyme polypeptide  
(R,K){20}  
238: SFANA KRRRRRRRRRRRRRRRRRRR KKKKG  
(K){20}  
239: FANAK KRRRRRRRRRRRRRRRRRRR KKKGR  
(K){20}  
240: ANAKK KRRRRRRRRRRRRRRRRRRR KKGRR  
(K){20}  
241: NAKKK KRRRRRRRRRRRRRRRRRRR KGRPX  
(K){20}  
242: AKKKK KRRRRRRRRRRRRRRRRRRR GRPXX  
(K){20}

1 AAU27944 ck: 1121 len: 69 i AAU27944 Human contig polypeptide sequen  
(R,K){20}  
40: VPPLT RRRRRRRRRRRRRRRRRRRR KKKKK  
(R,K){20}  
41: PPLTR KRRRRRRRRRRRRRRRRRRR KKKKK  
(K){20}  
42: PLTRK KRRRRRRRRRRRRRRRRRRR KKKKR  
(K){20}  
43: LTRKK KRRRRRRRRRRRRRRRRRRR KKKRG  
(K){20}  
44: TRKKK KRRRRRRRRRRRRRRRRRRR KKRGA  
(K){20}  
45: RKKKK KRRRRRRRRRRRRRRRRRRR KRGAL  
(K){20}  
46: KKKKK KRRRRRRRRRRRRRRRRRRR RGAL  
(K){20}  
47: KKKKK KRRRRRRRRRRRRRRRRRRR GAL  
(R,K){20}

1 AAU31467 ck: 4264 len: 657 i AAU31467 Novel human secreted protein #1  
(R,K){20}  
19: RRRRP RRRRRRRRRRRRRRRRRRRR RRRRL  
(R){20}  
20: RRRRP RRRRRRRRRRRRRRRRRRRR RRRLG  
(R){20}

21: RPRR RRRRRRRRRRRRRR RLGL  
(R){20}  
22: RPRR RRRRRRRRRRRRRR RLGL  
(R){20}  
23: PRRR RRRRRRRRRRRRRR LGLE

Aa03348 ck: 8085 len: 154 1 Aa03348 Human breast cancer protein encode  
(R,K){20}

114: TOLR KKKKKKKKKKKKKKKK KKKK  
(K){20}

115: OLRO KKKKKKKKKKKKKKKK KKKK  
(K){20}

116: LROK KKKKKKKKKKKKKKKK KKKK  
(K){20}

117: ROKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

118: OKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

119: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

120: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

121: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

122: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

123: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

124: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

125: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

126: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

127: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

128: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

1 Aa09664 ck: 1663 len: 87 1 Aa09664 Human pancreatic related protein H  
(R,K){20}

36: KWSS KKKKKKKKKKKKKKKK KKKK  
(K){20}

37: WSSX KKKKKKKKKKKKKKKK KKKK  
(K){20}

38: SSXX KKKKKKKKKKKKKKKK KKKK  
(K){20}

39: SXKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

40: XKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

41: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

42: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

43: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

44: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

45: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

46: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

47: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

48: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

49: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

50: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

51: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

52: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

53: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

54: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

55: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

56: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

57: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

58: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

59: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

60: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

61: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

62: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}

63: KKKK KKKKKKKKKKKKKKKK KKKK  
(R,K){20}

```
1 AAM95365 ck: 5626 len: 139 ! Aam95365 Human reproductive system related
      (R,K){20}
      (K){20}
111: IHLNL KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
112: HLNLK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
113: LNLKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
114: NLKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
AAM9607 ck: 4751 len: 80 ! Aam9607 Human reproductive system related
      (R,K){20}
      (K){20}
61: KXXFD KKKKKKKKKKKKKKKKKKK
      (K){20}
1 AAU18162 ck: 7907 len: 39 ! Aau18162 Novel human DNA-binding protein #9
      (R,K){20}
      (K){20}
9: YFEDL KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
10: FEDLK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
11: EDLKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
12: DLKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
13: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
1 AAU18167 ck: 9194 len: 87 ! Aau18167 Novel human DNA-binding protein #1
      (R,K){20}
      (K){20}
52: KIILL KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
53: IILLK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
54: ILKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
55: LLKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
56: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
1 AAU18168 ck: 8659 len: 104 ! Aau18168 Novel human DNA-binding protein #1
      (R,K){20}
      (K){20}
75: PLGGQ KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}

1 76: LGGQK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
77: GGQKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
78: GQKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
79: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
80: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
1 AAU18171 ck: 9398 len: 48 ! Aau18171 Novel human DNA-binding protein
      (R,K){20}
      (K){20}
2: Q KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
3: QK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
4: QKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
1 AAU18177 ck: 8278 len: 53 ! Aau18177 Novel human DNA-binding protein
      (R,K){20}
      (K){20}
30: NCGIL KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
31: CGILK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
32: GILKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
33: ILKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
34: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
1 AAU18178 ck: 444 len: 66 ! Aau18178 Novel human DNA-binding protein
      (R,K){20}
      (K){20}
35: SMTFS KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
36: MTFSS KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
1 AAU18179 ck: 5503 len: 50 ! Aau18179 Novel human DNA-binding protein
      (R,K){20}
      (K){20}
30: IICLL KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
31: ICCLK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
1 AAU18184 ck: 5691 len: 108 ! Aau18184 Novel human DNA-binding protein
      (R,K){20}
      (K){20}
78: VRPCL KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
```



(K){20}  
79: RPLK KKKKKKKKKKKKKKKKKKK  
(K){20}  
80: PCLK KKKKKKKKKKKKKKKKKKK  
(K){20}  
81: CLKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
82: LKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
83: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
84: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
85: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
86: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
87: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
88: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
89: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AAU18192 ck: 6029 len: 63 i Aau18192 Novel human DNA-binding protein #3  
(R,K){20}  
(K){20}  
40: KLTL KKKKKKKKKKKKKKKKKKK ISMG

1

AAU18200 ck: 7170 len: 63 i Aau18200 Novel human DNA-binding protein #4  
(R,K){20}  
(K){20}  
37: TPRS KKKKKKKKKKKKKKKKKKK  
(K){20}  
38: PSRA KKKKKKKKKKKKKKKKKKK  
(K){20}  
39: SRAK KKKKKKKKKKKKKKKKKKK  
(K){20}  
40: RAKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
41: AKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
42: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
43: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
44: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AAU18204 ck: 6110 len: 61 i Aau18204 Novel human DNA-binding protein #5  
(R,K){20}

(K){20}  
28: RPTRP KKKKKKKKKKKKKKKKKKK  
(K){20}  
29: PTPPK KKKKKKKKKKKKKKKKKKK  
(K){20}  
30: TRPK KKKKKKKKKKKKKKKKKKK  
(K){20}  
31: RPKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
32: PKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
33: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
34: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
35: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
36: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
37: KKKK KKKKKKKKKKKKKKKKKKK GGGG

AAU18205 ck: 5764 len: 74 i Aau18205 Novel human DNA-binding prote:  
(R,K){20}  
(K){20}  
40: EFLSA KKKKKKKKKKKKKKKKKKK  
(K){20}  
41: FLSAK KKKKKKKKKKKKKKKKKKK  
(K){20}  
42: LSAA KKKKKKKKKKKKKKKKKKK  
(K){20}  
43: SAKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
44: AKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
45: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
46: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
47: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
48: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
49: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
50: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
51: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
52: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKK KX

(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKK X

1  
AAU18206 ck: 9217 len: 68 i Aau18206 Novel human DNA-binding protein #5  
(R,K){20}

(K){20}  
38: FLFPE KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
39: LFPEK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
40: FPEKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
41: PEKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
42: EKKKK KKKKKKKKKKKKKKKKK KKKGX

(K){20}  
43: KKKKK KKKKKKKKKKKKKKKKK KKGXX

(K){20}  
44: KKKKK KKKKKKKKKKKKKKKKK KGXXF

(K){20}  
45: KKKKK KKKKKKKKKKKKKKKKK GXXF

1  
AAU18208 ck: 8162 len: 79 i Aau18208 Novel human DNA-binding protein #5  
(R,K){20}

(R,K){20}  
41: VRPRV RKKKKKKKKKKKKKKKK KKKKG

(K){20}  
42: RPRVR KKKKKKKKKKKKKKKKK KKKGG

(K){20}  
43: PRVRK KKKKKKKKKKKKKKKKK KKGGR

(K){20}  
44: RVRRK KKKKKKKKKKKKKKKKK KGGRF

(K){20}  
45: VRKKK KKKKKKKKKKKKKKKKK GGRFR

1  
AAU18237 ck: 285 len: 118 i Aau18237 Novel human DNA-binding protein #8  
(R,K){20}

(R,K){20}  
98: EKHKQ KKKKKKKKKKKKKKKKKR G

1  
AAU18238 ck: 5509 len: 58 i Aau18238 Novel human DNA-binding protein #8  
(R,K){20}

(K){20}  
36: FYFVC KKKKKKKKKKKKKKKKK KKK

(K){20}  
37: YFVCK KKKKKKKKKKKKKKKKK KK

(K){20}  
38: FVCKK KKKKKKKKKKKKKKKKK K

(K){20}  
39: VCKKK KKKKKKKKKKKKKKKKK

1  
AAU18239 ck: 9074 len: 66 i Aau18239 Novel human DNA-binding protein  
(R,K){20}

(K){20}  
40: LVOCE KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
41: VOCEK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
42: QCEKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
43: CEKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
44: EKKKK KKKKKKKKKKKKKKKKK KKK

(K){20}  
45: KKKKK KKKKKKKKKKKKKKKKK KK

(K){20}  
46: KKKKK KKKKKKKKKKKKKKKKK K

(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKK

1  
AAU18240 ck: 8528 len: 150 i Aau18240 Novel human DNA-binding protein  
(R,K){20}

(K){20}  
113: SRNTV KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
114: RNTVK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
115: NTVRK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
116: TVRRK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
117: VRKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
118: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
119: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
120: KKKKK KKKKKKKKKKKKKKKKK KKKXX

(K){20}  
121: KKKKK KKKKKKKKKKKKKKKKK KKKXX

(K){20}  
122: KKKKK KKKKKKKKKKKKKKKKK KXXAV

(K){20}  
123: KKKKK KKKKKKKKKKKKKKKKK XHAVL

1  
AAU18241 ck: 7676 len: 156 i Aau18241 Novel human DNA-binding protein  
(R,K){20}

(K){20}

108: KTTWI KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
109: TTWIK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
110: TWIKR KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
111: WIKKK KKKKKKKKKKKKKKKKKKK KKKRG  
(K){20}  
112: IKKKK KKKKKKKKKKKKKKKKKKK KKRGX  
(K){20}  
113: KKKKK KKKKKKKKKKKKKKKKKKK KRGXG  
(K){20}  
114: KKKKK KKKKKKKKKKKKKKKKKKK RGXGS  
(R,K){20}  
115: KKKKK KKKKKKKKKKKKKKKKKR GXGST

AAU18242 ck: 1736 len: 40 ! Aau18242 Novel human DNA-binding protein #8  
(R,K){20}  
18: LPGSL KKKKKKKKKKKKKKKKKKK KKK  
(K){20}  
19: PGSLL KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
20: GSLKK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
21: SLKKK KKKKKKKKKKKKKKKKKKK

AAU18244 ck: 1109 len: 98 ! Aau18244 Novel human DNA-binding protein #9  
(R,K){20}  
53: QTKNT KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
54: TKNTK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
55: KNTKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
56: NTKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
57: TKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

(K){20}  
63: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
64: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
65: KKKKK KKKKKKKKKKKKKKKKKKK KKKGG  
(K){20}  
66: KKKKK KKKKKKKKKKKKKKKKKKK KKGRR  
(K){20}  
67: KKKKK KKKKKKKKKKKKKKKKKKK KGGRR  
(K){20}  
68: KKKKK KKKKKKKKKKKKKKKKKKK GGRSR

AAU18246 ck: 8102 len: 111 ! Aau18246 Novel human DNA-binding prote  
(R,K){20}  
78: EFHIL KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
79: FHILK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
80: HILKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
81: ILKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
82: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
83: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
84: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
85: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
86: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
87: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
88: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
89: KKKKK KKKKKKKKKKKKKKKKKKK KKK  
(K){20}  
90: KKKKK KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
91: KKKKK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
92: KKKKK KKKKKKKKKKKKKKKKKKK

AAU18247 ck: 8102 len: 111 ! Aau18247 Novel human DNA-binding protei  
(R,K){20}  
78: EFHIL KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

1

79: FHILK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

80: HILKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

81: ILKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

82: LKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

83: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

84: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

85: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

86: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

87: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

88: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

89: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

90: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

91: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

92: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

AAU18248 ck: 8319 len: 53 1 Aau18248 Novel human DNA-binding protein #9

(R,K)120  
(K)120

13: RYFKP KKKKKKKKKKKKKKKKKKKKKKKKK

14: YFKPK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

15: FKPKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

16: KPKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

17: PKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

18: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

19: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

20: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

21: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

1

22: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

23: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

24: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

25: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

26: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

27: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

28: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

29: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

30: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

AAU18250 ck: 7918 len: 80 1 Aau18250 Novel human DNA-binding protein

(R,K)120  
(K)120

50: NVLTV KKKKKKKKKKKKKKKKKKKKKKKKK

51: VLTVK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

52: LTVKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

53: TVKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

54: VKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

55: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

56: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

AAU18252 ck: 4882 len: 41 1 Aau18252 Novel human DNA-binding protein

(R,K)120  
(K)120

8: FYCFP KKKKKKKKKKKKKKKKKKKKKKKKK

9: YCFPK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

10: CFPKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

11: FFKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

12: FKKKK (K)120  
KKKKKKKKKKKKKKKKKKKKKKKKKK

13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKK GX  
(K){20}

AAU18253 ck: 5469 len: 63 i Aau18253 Novel human DNA-binding protein #1

1

(R,K){20}  
(K){20}  
30: IICLL KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
31: ICLLK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
32: CLKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
33: LLKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
34: LKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
35: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
36: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
37: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
38: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
39: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
40: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
41: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
42: KKKK KKKKKKKKKKKKKKKKKKK XG  
(K){20}

AAU18254 ck: 5075 len: 52 i Aau18254 Novel human DNA-binding protein #1

1

(R,K){20}  
(K){20}  
30: FIVVX KKKKKKKKKKKKKKKKKKK KKK

(K){20}  
31: IVVX KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
32: VVXX KKKKKKKKKKKKKKKKKKK K  
(K){20}  
33: VXXX KKKKKKKKKKKKKKKKKKK

AAU18255 ck: 5741 len: 47 i Aau18255 Novel human DNA-binding prote

1

(R,K){20}  
(K){20}  
20: ILTF KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: LTFK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: TFKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: TFKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: FKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
27: KKKK KKKKKKKKKKKKKKKKKKK X  
(K){20}

AAU18256 ck: 2868 len: 84 i Aau18256 Novel human DNA-binding prote

1

(R,K){20}  
(K){20}  
53: KCTE KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
54: CTYE KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
55: TYEK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
56: YEKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
57: EKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
58: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
59: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
60: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
61: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
62: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AAU18257 ck: 4686 len: 73 i Aau18257 Novel human DNA-binding prote

1

(R,K){20}  
 41: YLKEE KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 42: LKKEE KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 43: KKEKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 44: KEKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 45: EKKEK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 51: KKKKK KKKKKKKKKKKKKKKKKKK KKK  
 (K){20}  
 52: KKKKK KKKKKKKKKKKKKKKKKKK KK  
 (K){20}  
 53: KKKKK KKKKKKKKKKKKKKKKKKK K  
 (K){20}  
 54: KKKKK KKKKKKKKKKKKKKKKKKK

1

AAU18258 ck: 6676 len: 74 i Aau18258 Novel human DNA-binding protein #1  
 (R,K){20}  
 47: LRTFQ KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 48: RTFQK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 49: TFOKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 50: FOKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 51: OKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}

1

AAU18259 ck: 2283 len: 54 i Aau18259 Novel human DNA-binding protein #1  
 (R,K){20}  
 32: IVFCF KKKKKKKKKKKKKKKKKKK KKK

1

(K){20}  
 33: VFCEK KKKKKKKKKKKKKKKKKKK KX  
 (K){20}  
 34: FCEKK KKKKKKKKKKKKKKKKKKK X

AAU18260 ck: 7503 len: 74 i Aau18260 Novel human DNA-binding protein  
 (R,K){20}  
 45: SHLTD KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 46: HLTDK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 47: LTDKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 48: TDKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 49: DKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 52: KKKKK KKKKKKKKKKKKKKKKKKK KKK  
 (K){20}  
 53: KKKKK KKKKKKKKKKKKKKKKKKK KK  
 (K){20}  
 54: KKKKK KKKKKKKKKKKKKKKKKKK K  
 (K){20}  
 55: KKKKK KKKKKKKKKKKKKKKKKKK

1

AAU18262 ck: 5199 len: 84 i Aau18262 Novel human DNA-binding protein  
 (R,K){20}  
 63: AMNAS KKKKKKKKKKKKKKKKKKK XG  
 (K){20}

1

AAU18263 ck: 7578 len: 31 i Aau18263 Novel human DNA-binding protein  
 (R,K){20}  
 6: LTELK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 7: TELEK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 8: ELEKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 9: LEKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 10: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 11: KKKKK KKKKKKKKKKKKKKKKKKK X

```
1
AAU18264 ck: 3915 len: 57 i Aau18264 Novel human DNA-binding protein #1
(R,K){20}
31: KQLLK KKKKKKKKKKKKKKKKKKKKK KKKXG
(K){20}
32: QLLLK KKKKKKKKKKKKKKKKKKKKK KXXGG
(K){20}
33: LLLK KKKKKKKKKKKKKKKKKKKKK KXGGF
(K){20}
34: LLKK KKKKKKKKKKKKKKKKKKKKK XGGF
(K){20}

AAU18265 ck: 3679 len: 37 i Aau18265 Novel human DNA-binding protein #1
(R,K){20}
15: ISPLT KKKKKKKKKKKKKKKKKKKKK KXX
(K){20}
16: SPLTK KKKKKKKKKKKKKKKKKKKKK KX
(K){20}
17: PLTK KKKKKKKKKKKKKKKKKKKKK X
(K){20}

AAU18266 ck: 657 len: 196 i Aau18266 Novel human DNA-binding protein #1
(R,K){20}
169: FVXFE KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
170: VXFEX KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
171: XFEX KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
172: FEXKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
173: EXKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
174: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}
175: KKKKK KKKKKKKKKKKKKKKKKKKKK KX
(K){20}
176: KKKKK KKKKKKKKKKKKKKKKKKKKK X
(K){20}

AAU18267 ck: 4672 len: 57 i Aau18267 Novel human DNA-binding protein #1
(R,K){20}
28: DKTFH KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: KTFHK KKKKKKKKKKKKKKKKKKKKK KKKXP
(K){20}
30: TFHKK KKKKKKKKKKKKKKKKKKKKK KKKXP
(K){20}
31: FHKKK KKKKKKKKKKKKKKKKKKKKK KXPGG
(K){20}

1
32: HKKKK KKKKKKKKKKKKKKKKKKKKK XPGCG
(K){20}
AAU18268 ck: 9656 len: 66 i Aau18268 Novel human DNA-binding protein #1
(R,K){20}
38: MVISV KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
39: VISVK KKKKKKKKKKKKKKKKKKKKK KKKRE
(K){20}
40: ISVKK KKKKKKKKKKKKKKKKKKKKK KKKREK
(K){20}
41: SVKKK KKKKKKKKKKKKKKKKKKKKK KKKREX
(K){20}
42: VKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

AAU18270 ck: 4665 len: 34 i Aau18270 Novel human DNA-binding protein #1
(R,K){20}
10: PELLK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: ELLK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: LLKK KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}
13: LLKK KKKKKKKKKKKKKKKKKKKKK KK
(K){20}
14: LKKK KKKKKKKKKKKKKKKKKKKKK K
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}

AAU18271 ck: 7810 len: 64 i Aau18271 Novel human DNA-binding protein #1
(R,K){20}
37: LKYFW KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
38: KYFWK KKKKKKKKKKKKKKKKKKKKK KKKXG
(K){20}
39: YFWKK KKKKKKKKKKKKKKKKKKKKK KKKXG
(K){20}
40: FWKKK KKKKKKKKKKKKKKKKKKKKK KXGXP
(K){20}
41: WKKKK KKKKKKKKKKKKKKKKKKKKK XGXP
(K){20}

AAU18272 ck: 269 len: 76 i Aau18272 Novel human DNA-binding protein #1
(R,K){20}
37: TFSRA KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
```

38: PSRAK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
39: SHAKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
40: RAKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
41: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

AAU18273 ck: 8370 len: 45 ! Aau18273 Novel human DNA-binding protein #1

(R,K) (20)  
17: APRTQ KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
18: PKTOK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
19: KTOKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
20: TOKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
21: OKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

AAU18274 ck: 1663 len: 87 ! Aau18274 Novel human DNA-binding protein #1

(R,K) (20)  
36: KWSXX KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
37: WSSXX KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
38: SSXXX KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
39: SXXXX KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
40: XXXXX KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)  
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
63: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
64: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

(K) (20)  
65: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)

AAU18275 ck: 5607 len: 63 ! Aau18275 Novel human DNA-binding protein

(R,K) (20)  
26: MVELE KKKKKKKKKKKKKKKKKKK KKKKK  
(K) (20)



27: VLEK (K)120  
KKKKKKKKKKKKKKKKKKKK  
28: ELEK (K)120  
KKKKKKKKKKKKKKKKKKKK  
29: LEKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
30: EKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
31: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
32: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
33: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
34: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
35: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
36: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
37: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
38: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
39: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
40: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
AAU18276 ck: 5997 len: 58 i Aau18276 Novel human DNA-binding protein #1  
(R,K)120  
(K)120  
28: RPTR KKKKKKKKKKKKKKKKKKKKK  
29: PTRPK KKKKKKKKKKKKKKKKKKKKK  
AAU18277 ck: 5764 len: 74 i Aau18277 Novel human DNA-binding protein #1  
(R,K)120  
(K)120  
40: EPLSA KKKKKKKKKKKKKKKKKKKKK  
41: FLSAK KKKKKKKKKKKKKKKKKKKKK  
42: LSARK KKKKKKKKKKKKKKKKKKKKK  
43: SAKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
44: AKKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
45: KKKKK (K)120  
KKKKKKKKKKKKKKKKKKKK

46: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
47: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
48: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
49: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
50: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
51: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
52: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
53: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
54: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
AAU18278 ck: 7734 len: 97 i Aau18278 Novel human DNA-binding protein #1  
(R,K)120  
(K)120  
57: RGCSY KKKKKKKKKKKKKKKKKKKKK  
58: GCSYK KKKKKKKKKKKKKKKKKKKKK  
59: CSYKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
60: SYKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
61: YKKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
62: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
63: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
64: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
65: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
66: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
67: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
68: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
69: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK  
70: KKKK (K)120  
KKKKKKKKKKKKKKKKKKKK

71: KKKKK (K){20} KKKKK  
72: KKKKK (K){20} KKKKK  
73: KKKKK (K){20} KKKKK  
74: KKKKK (K){20} KKKKK  
75: KKKKK (K){20} KKKKK

AAU18279 ck: 3164 len: 181 i Aau18279 Novel human DNA-binding protein #1  
(R,K){20}  
154: TRKPE KKKKKKKKKKKKKKKKKKK  
155: RKPEK (K){20} KKKKK  
156: KPEKK (K){20} KKKKK  
157: PEKKK (K){20} KKKKK  
158: EKKKK (K){20} KKKK  
159: KKKKK (K){20} KKKK  
160: KKKKK (K){20} KK  
161: KKKKK (K){20} K

1  
1  
AAU18280 ck: 7117 len: 55 i Aau18280 Novel human DNA-binding protein #1  
(R,K){20}  
22: DDKNN KKKKKKKKKKKKKKKKKKK  
23: DKKNK (K){20} KKKKK  
24: KKNKK (K){20} KKKKK  
25: KNNKK (K){20} KKKKK  
26: NKKKK (K){20} KKKKK  
27: KKKKK (K){20} KKKKK  
AAU18281 ck: 9316 len: 67 i Aau18281 Novel human DNA-binding protein #1  
(R,K){20}  
36: FTLOT RKKKKKKKKKKKKKKKKKK

37: TLOTR (K){20} KKKKK  
38: LOTRK (K){20} KKKKK  
39: QTRKK (K){20} KKKKK  
40: TRKKK (K){20} KKKKK  
41: RKKKK (K){20} KKKKK

AAU53862 ck: 5383 len: 86 i Aau53862 Human brain expressed single ex  
(R,K){20}  
15: RRRRG RRRRRKKKKKKKKKKKKKK  
16: RRRGR RRRRRKKKKKKKKKKKKKK  
17: RRGRR RRRRRKKKKKKKKKKKKKK  
18: RGRRR RRRRRKKKKKKKKKKKKKK  
19: GRRRR RRRRRKKKKKKKKKKKKKK  
20: RRRRR (R,K){20} KKKKK  
21: RRRRK (R,K){20} KKKKK  
22: RRRKK (R,K){20} KKKKK  
23: RRRKK (R,K){20} KKKKK  
24: RKKKK (R,K){20} KKKKK  
25: KKKKK (R,K){20} KKKKK  
26: KKKKK (R,K){20} KKKKK  
27: KKKKR (R,K){20} KKKKK  
28: KKKRK (R,K){20} KKKKK  
29: KKKKK (R,K){20} KKKKK  
30: KKKKK (R,K){20} KKKKK  
31: RKKKK (R,K){20} KKKKK  
32: KKKKK (R,K){20} KKKKK

(R,K){20}

33: KKKK KKKRRRRRRRRRRRRRR RR

(R,K){20}

34: KKKK KKKRRRRRRRRRRRRRR RR

(R,K){20}

35: KKKK KKKRRRRRRRRRRRRRR RR

(R,K){20}

36: KKKK KRRRRRRRRRRRRRRRR RR

(R){20}

37: KKKK RRRRRRRRRRRRRRRRR RR

(R){20}

38: KKKK RRRRRRRRRRRRRRRRR RR

(R){20}

39: KKKK RRRRRRRRRRRRRRRRR RR

(R){20}

40: KKKK RRRRRRRRRRRRRRRRR RR

(R){20}

41: KRRR RRRRRRRRRRRRRRRRR RR

(R){20}

42: RRRR RRRRRRRRRRRRRRRRR RR

(R){20}

43: RRRR RRRRRRRRRRRRRRRRR RR

(R){20}

44: RRRR RRRRRRRRRRRRRRRRR RR

(R){20}

45: RRRR RRRRRRRRRRRRRRRRR RR

AAM54700 ck: 1334 len: 86 i Aam54700 Human brain expressed single exon

```

(R,K){20}
(R,K){20}
57: EEEEG RRRKKKKRKKKKKKKKKKK KKKK
      (R,K){20}
58: EEGR RRRKKKKRKKKKKKKKKKK KKKK
      (R,K){20}
59: EEGR RKKKKRKKKKKKKKKKKK KKKK
      (R,K){20}
60: EGRR KKKKKRKKKKKKKKKKKK KKKK
      (R,K){20}
61: GRRR KKKKRKKKKKKKKKKKK KKKK
      (R,K){20}
62: RRRK KKKRKKKKKKKKKKKKKK KKKK
      (R,K){20}
63: RRKK KRKKKKKKKKKKKKKKKK KKKK
      (R,K){20}
64: RKKK RRRKKKKKKKKKKKKKK KK
      (R,K){20}
65: KKKR KRKKKKKKKKKKKKKK KK
      (R,K){20}
66: KKKR RKKKKKKKKKKKKKKKK K

```

```

67 : KKKRR KKKKKKKKKKKKKKKKKK      (K){20}

AAM54795   ck: 9082   len: 167   ! Aam54795 Human brain expressed single e

33 : EEGRG RRRRRRRRRRRRRRRRKRKRGCG    (R,K){20}
        (R,K){20}

34 : EGRGR RRRRRRRRRRRRRRRRKRKGGGR     (R,K){20}
        (R,K){20}

35 : GGRGR RRRRRRRRRRRRRRRRKRKGGGR

```

```

AAM55320 ck: 2276 len: 89 1 Aam55320 Human brain expressed single e
(R,K)[20]
23: EEEEE KKKKKKKKKKKKKKKKKKK EEEKK
(R,K)[20]
46: KKEEE KKKKKKKKKKKKKKKKKKK KKEEE
(K)[20]
47: KEEKK KKKKKKKKKKKKKKKKKKK KKEEE
(K)[20]
48: EEEKK KKKKKKKKKKKKKKKKKKK KEEEE
(R,K)[20]
49: EEKKK KKKKKKKKKKKKKKKKKKK EEEEE
(R,K)[20]
AAM55623 ck: 1939 len: 130 1 Aam55623 Human brain expressed single e
(R,K)[20]
42: EGKKE RRRRRRRRRRRRRRRRRRR RRRRR
(R,K)[20]
43: GKKEK RRRRRRRRRRRRRRRRRRR RRRRR
(R,K)[20]
44: RKEER RRRRRRRRRRRRRRRRRRR RRRRK
(R,K)[20]
45: KEERR RRRRRRRRRRRRRRRRRRR RRRKK
(R,K)[20]
46: ERRRR RRRRRRRRRRRRRRRRRRR RRRKK
(R,K)[20]
47: RRRRR RRRRRRRRRRRRRRRRRRR RKKKK
(R,K)[20]
48: RRRRR RRRRRRRRRRRRRRRRRRR RKKKK
(R,K)[20]
49: RRRRR RRRRRRRRRRRRRRRRRRR RKKKK
(R,K)[20]
50: RRRRR RRRRRRRRRRRRRRRRRRR RKKKE
(R,K)[20]
51: RRRRR RRRRRRRRRRRRRRRRRRR RKKKE
(R,K)[20]
52: RRRRR RRRRRRRRRRRRRRRRRRR KKEEE
(R,K)[20]

```

53: RRRRR RRRRRRRRRRRRRRRRRRRR KEEEE  
(R,K){20}  
54: RRRRR RRRRRRRRRRRRRRRRRRRR EEEEE  
AAM55931 ck: 3607 len: 88 i Aam55931 Human brain expressed single exon  
(R,K){20}  
39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
40: RKRER KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
41: KREKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
42: REKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
67: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
68: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
69: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AAM56490 ck: 3937 len: 85 i Aam56490 Human brain expressed single exon  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AAM58224 ck: 1560 len: 88 i Aam58224 Human brain expressed single exon  
(R,K){20}  
43: RERRG RRRRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}

[illegible]

```

7: TEREK KKKKKKKKKKKKKKKKKK KNKK      (R,K){20}
8: EREKK KKRKKKKKKKKKKKKKKKK KNKK      (R,K){20}
9: REKKK KRKKKKKKKKKKKKKKKK NKKK       (R,K){20}

AAM63533 ck: 2394 len: 57 i Aam63533 Human brain expressed single e
(R,K){20}
20: EEDEG RRRRRRRRRRRRRRRRRR RRRR      (R,K){20}
21: EEGGR RRRRRRRRRRRRRRRRRR RRRG      (R,K){20}
22: EGGR RRRRRRRRRRRRRRRRRR RRGR      (R,K){20}
23: EGRR RRRRRRRRRRRRRRRRRR RRGR      (R,K){20}
24: GRRR RRRRRRRRRRRRRRRRRR RGRR      (R,K){20}
25: RRRR RRRRRRRRRRRRRRRRRR GRRR      (R,K){20}

AAM64090 ck: 4228 len: 24 i Aam64090 Human brain expressed single e
(R,K){20}
1: RRRRRRRRRRRRRRRRRR RKKT            (R,K){20}
2: R RRRRRRRRRRRRRRRR KRT             (R,K){20}
3: RR RRRRRRRRRRRRRRRR RT              (R,K){20}
4: RRR RRRRRRRRRRRRRRRR T              (R,K){20}

AAM64863 ck: 3301 len: 52 i Aam64863 Human brain expressed single e
(R,K){20}
12: KKKKN KKKKKKKKKRRKKKKKK KKKK      (R,K){20}
13: KKKNK KKKKKKKRRKKKKKK KKKK        (R,K){20}
14: KKKNK KKKKKRRKKKKKKKK KKKK        (R,K){20}
15: KKKKK KKKKKRRKKKKKKKK KKKK        (R,K){20}
16: NKKKK KKKKKRRKKKKKKKK KKKK        (R,K){20}
17: KKKKK KKKKKRRKKKKKKKK KKKK        (R,K){20}
18: KKKKK KKKKKRRKKKKKKKK KKKK        (R,K){20}

```

```

(R,K){20}
19: KKKK KKKRKRKKKKKKKKKK KKKK
      (R,K){20}
20: KKKK KRKRKRKKKKKKKKKK KKKK
      (R,K){20}
21: KKKK KRKRKRKKKKKKKKKK KKKK
      (R,K){20}
22: KKKR KRKRKKKKKKKKKKKK KKKK
      (R,K){20}
23: KKKR KRKRKKKKKKKKKKKK KKKK
      (R,K){20}
24: KKKR KRKRKKKKKKKKKKKK KKKK
      (R,K){20}
25: KRRK RRRKKKKKKKKKKKKKK KKKK
      (R,K){20}
26: RRRK RRRKKKKKKKKKKKKKK KKKK
      (R,K){20}
27: KRRK RRRKKKKKKKKKKKKKK KKKK
      (K){20}
28: RRRR RRRKKKKKKKKKKKKKK KKAF
      (K){20}
29: KRRK RRRKKKKKKKKKKKKKK KKAF
      (K){20}
30: RRRK RRRKKKKKKKKKKKKKK KA
      (K){20}
31: KRRK RRRKKKKKKKKKKKKKK AF

AAM6535 ck: 4895 len: 51 | Aam6535 Human brain expressed single exon
      (R,K){20}
18: LFRPM RRRRRKKRRRRKKRRR KLTT
      (R,K){20}
19: FRPR KRKRKKRRRRKKRRR KLTT
      (R,K){20}
20: KPRK RRRKKRRRRKKRRR LTTT

AAM6249 ck: 5383 len: 86 | Aam6249 Human bone marrow expressed probe
      (R,K){20}
15: RRRG RRRRRKKRRRRKKRRR KRRR
      (R,K){20}
16: RRGR RRRKKKKKKKKKKKKR KRRR
      (R,K){20}
17: RGRR RRRKKKKKKKKKKKKR RRRR
      (R,K){20}
18: RGRR RRRKKKKKKKKKKKKR RRRR
      (R,K){20}
19: GRRR RRRKKKKKKKKKKKKR RRRR
      (R,K){20}
20: RRRR RRRKKKKKKKKKKKKR RRRR

```

21:	RRRR	KKKKKKKKKKKKKKRRRR	RRRRR	(R,K){20}
22:	RRRR	KKKKKKKKKKKKKKRRRR	RRRRR	(R,K){20}
23:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
24:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
25:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
26:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
27:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
28:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
29:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
30:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
31:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
32:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
33:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
34:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
35:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
36:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R,K){20}
37:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R){20}
38:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R){20}
39:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R){20}
40:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R){20}
41:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R){20}
42:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R){20}
43:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R){20}
44:	RRRR	KKKKKKKKKKKKKKRRRRR	RRRRR	(R){20}

45: RRRRR (R){20}  
RRRRRRRRRRRRRRRRRRR NKQTK

AAM67100 ck: 1334 len: 86 ! Aam67100 Human bone marrow expressed probe

57: EEEEG (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKKK

58: EEEGR (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKKK

59: EEEGR (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKKK

60: EGRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKKK

61: GRRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKKK

62: RRRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKKK

63: RRRKK (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKK

64: RRRKK (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKK

65: KRRKK (R,K){20}  
RRRRRRRRRRRRRRRRRRR KK

66: KRRKK (R,K){20}  
RRRRRRRRRRRRRRRRRRR K

67: KRRRR (K){20}  
RRRRRRRRRRRRRRRRRRR

AAM67180 ck: 9082 len: 167 ! Aam67180 Human bone marrow expressed probe

33: EEEGR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRGCG

34: EEEGR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRGCG

35: GGGRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR GGGRR

AAM67717 ck: 2276 len: 89 ! Aam67717 Human bone marrow expressed probe

23: EEEEE (R,K){20}  
RRRRRRRRRRRRRRRRRRR EEEKK

46: KKEEE (K){20}  
RRRRRRRRRRRRRRRRRRR KKEEE

47: KKEEK (K){20}  
RRRRRRRRRRRRRRRRRRR KKEEE

48: EEEKK (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKEEE

49: EEEKK (R,K){20}  
RRRRRRRRRRRRRRRRRRR EEEEE

1

AAM68007 ck: 1939 len: 130 ! Aam68007 Human bone marrow expressed pr

42: EGRKE (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

43: GRRER (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

44: RKEER (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

45: KERER (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

46: ERERR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

47: RRRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

48: RRRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

49: RRRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

50: RRRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

51: RRRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

52: RRRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

53: RRRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

54: RRRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

AAM68298 ck: 3607 len: 88 ! Aam68298 Human bone marrow expressed pr

39: ERKRE (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

40: RREKK (K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

41: KREKK (K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

42: REKKK (K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

43: EKKKK (K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

44: KKKKK (K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

45: KKKKK (K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

46: KKKKK (K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

47: KKKKK (K){20}  
RRRRRRRRRRRRRRRRRRR RRRRR

48: kkkk (K){20} kkkk  
49: kkkk (K){20} kkkk  
50: kkkk (K){20} kkkk  
51: kkkk (K){20} kkkk  
52: kkkk (K){20} kkkk  
53: kkkk (K){20} kkkk  
54: kkkk (K){20} kkkk  
55: kkkk (K){20} kkkk  
56: kkkk (K){20} kkkk  
57: kkkk (K){20} kkkk  
58: kkkk (K){20} kkkk  
59: kkkk (K){20} kkkk  
60: kkkk (K){20} kkkk  
61: kkkk (K){20} kkkk  
62: kkkk (K){20} kkkk  
63: kkkk (K){20} kkkk  
64: kkkk (K){20} kkkk  
65: kkkk (K){20} kkkk  
66: kkkk (K){20} kkkk  
67: kkkk (K){20} kkkk  
68: kkkk (K){20} k  
69: kkkk (K){20} kkkk  
AAm68869 ck: 3937 len: 85 i Aam68869 Human bone marrow expressed probe  
(R,K){20}  
1: kkkkkkkkkkkkkkkkkk kkkk

2: k (K){20} kkkk  
3: k (K){20} kkkk  
4: k (K){20} kkkk  
5: k (K){20} kkkk  
6: k (K){20} kkkk  
7: k (K){20} kkkk  
8: k (K){20} kkkk  
9: k (K){20} kkkk  
10: k (K){20} kkkk  
11: k (K){20} kkkk  
AAm69569 ck: 2686 len: 71 i Aam69569 Human bone marrow expressed pro.  
(R,K){20}  
20: kkkk (K){20} kkkk  
21: kkkk (K){20} kkkk  
22: kkkk (K){20} kkkk  
23: kkkk (K){20} kkkk  
24: kkkk (K){20} kkkk  
25: kkkk (R,K){20} kkkk  
26: kkkk (R,K){20} kkkk  
27: kkkk (R,K){20} kkkk  
28: kkkk (R,K){20} kkkk  
29: kkkk (R,K){20} kkkk  
30: kkkk (R,K){20} kkkk  
31: kkkk (R,K){20} kkkk  
32: kkkk (R,K){20} kkkk



```
(R,K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
AAM70678 ck: 1560 len: 88 i Aam70678 Human bone marrow expressed probe
(R,K){20}
43: RRRRG RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
44: ERGRG RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
45: RRGRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
46: RGRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
47: GRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
48: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
```

```
(R){20}
49: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
50: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
51: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
52: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
53: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
54: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
55: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
56: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
57: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
58: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
59: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
60: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
61: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
62: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
AAM70881 ck: 2324 len: 36 i Aam70881 Human bone marrow expressed pr
(R,K){20}
16: KERKT KKKKKRRRRRRRRRRRRR R
(R,K){20}
17: ERKTK KKKKKRRRRRRRRRRRRR R
(R,K){20}
AAM73767 ck: 8343 len: 66 i Aam73767 Human bone marrow expressed pr
(R,K){20}
6: ETERE KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
7: TEREK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
8: EREKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
9: REKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
AAM76347 ck: 2394 len: 57 i Aam76347 Human bone marrow expressed pr
(R,K){20}
```

20: EEEEG RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
21: EEGGR RRRRRRRRRRRRRRRRRRR RRRRG  
(R){20}  
22: EEGRR RRRRRRRRRRRRRRRRRRR RRRGR  
(R){20}  
23: EGRRR RRRRRRRRRRRRRRRRRRR RRGRR  
(R){20}  
24: GRRRR RRRRRRRRRRRRRRRRRRR RGRRR  
(R){20}  
25: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}

1  
AAM76911 ck: 4228 len: 24 i Aam76911 Human bone marrow expressed probe  
(R,K){20}  
1: RRRRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
2: R RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
3: RR RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
4: RRR RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}

1  
AAM78048 ck: 4895 len: 51 i Aam78048 Human bone marrow expressed probe  
(R,K){20}  
18: LEPKM RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
19: EKPMM RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
20: KPMRK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}

1  
AAM82533 ck: 1736 len: 40 i Aam82533 Human immune/haematopoietic antigen  
(R,K){20}  
18: LPSGL RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
19: PPSGL RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
20: GSLKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
21: SLKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}

1  
AAM85748 ck: 7503 len: 74 i Aam85748 Human immune/haematopoietic antigen  
(R,K){20}  
45: SHLTD RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
46: HLTDK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}

47: LTDKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
48: TDKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
49: DKKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
50: KKKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
51: KKKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
52: KKKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
53: KKKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
54: KKKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
55: KKKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}

1  
AAM90546 ck: 6676 len: 74 i Aam90546 Human immune/haematopoietic antigen  
(R,K){20}  
47: LPTFQ RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
48: RPTFQ RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
49: TPTFQ RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
50: FQKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
51: QKKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
52: KKKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}

1  
AAM90618 ck: 5691 len: 108 i Aam90618 Human immune/haematopoietic antigen  
(R,K){20}  
78: VRPCL RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
79: RPLCK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
80: PCLKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
81: CLKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
82: LKKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
83: KKKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
84: KKKKK RRRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}

85: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
86: kkkk (K){20} kkkkkkkkkkkkkkkkk kkk  
87: kkkk (K){20} kkkkkkkkkkkkkkkkk k  
88: kkkk (K){20} kkkkkkkkkkkkkkkkk k  
89: kkkk (K){20} kkkkkkkkkkkkkkkkk

1  
AAM91030 ck: 8102 len: 111 ! Aam91030 Human immune/haematopoietic antige  
(R,K){20}

78: EFHIL kkkkkkkkkkkkkkkkkkk kkkk  
79: FHILK kkkkkkkkkkkkkkkkkkk kkkk  
80: HILKk kkkkkkkkkkkkkkkkkkk kkkk  
81: ILKkK kkkkkkkkkkkkkkkkkkk kkkk  
82: LKKkK kkkkkkkkkkkkkkkkkkk kkkk  
83: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
84: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
85: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
86: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
87: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
88: kkkk (K){20} kkkkkkkkkkkkkkkkk kkk  
89: kkkk (K){20} kkkkkkkkkkkkkkkkk k  
90: kkkk (K){20} kkkkkkkkkkkkkkkkk k  
91: kkkk (K){20} kkkkkkkkkkkkkkkkk k  
92: kkkk (K){20} kkkkkkkkkkkkkkkkk

1  
AAM91162 ck: 1109 len: 98 ! Aam91162 Human immune/haematopoietic antige  
(R,K){20}

53: QTKNT kkkkkkkkkkkkkkkkkkk kkkk  
54: TKNTK kkkkkkkkkkkkkkkkkkk kkkk

1

55: kNTKK (K){20} kkkkkkkkkkkkkkkkk kkkk  
56: NTKKK (K){20} kkkkkkkkkkkkkkkkk kkkk  
57: TKKKK (K){20} kkkkkkkkkkkkkkkkk kkkk  
58: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
59: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
60: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
61: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
62: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
63: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
64: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
65: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
66: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
67: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
68: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk

AAM91891 ck: 8102 len: 111 ! Aam91891 Human immune/haematopoietic an  
(R,K){20}

78: EFHIL kkkkkkkkkkkkkkkkkkk kkkk  
79: FHILK kkkkkkkkkkkkkkkkkkk kkkk  
80: HILKk kkkkkkkkkkkkkkkkkkk kkkk  
81: ILKkK kkkkkkkkkkkkkkkkkkk kkkk  
82: LKKkK kkkkkkkkkkkkkkkkkkk kkkk  
83: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
84: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
85: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk  
86: kkkk (K){20} kkkkkkkkkkkkkkkkk kkkk

[illegible]

```

1
AAM9193  CR: 8285  len: 55  ! Aam9193  Human digestive system antigen SEQ
      (R,K){20}
33:  PTRRP  KKKKKKKKKKKKKKKKKKKKK  KCK
      (K){20}
34:  PTRRP  KKKKKKKKKKKKKKKKKKKKK  CK

```

1  
AAM92433 cr: 1663 len: 87 | Aam92433 Human digestive system antigen SEC

(R,K){20}  
36: KWSXK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
37: WSSXK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
38: SSXK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
39: SXK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
40: XKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
41: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
42: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
43: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
44: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
45: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
46: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
47: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
48: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

```

49:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
50:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
51:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
52:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
53:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
54:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
55:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
56:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
57:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
58:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
59:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
60:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
61:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
62:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)
63:  KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) (20)

AA000092 ck: 9065 len: 113 i Aa000092 Human polypeptide seq ID NO 139
      (R,K) (20)
      (K) (20)
91:  CLGCL KKKKKKKKKKKKKKKKKKK KFF
      (K) (20)
92:  LGCLK KKKKKKKKKKKKKKKKKKK FF
      (K) (20)

```

```

1
AA000222  CK: 2916  len: 132  ! Aa00222 Human polypeptide SEQ ID NO 141
      (R,K){20}
28: XLPKP KKKKKKKKKKKKKKKKKKKKK GSPPK
      (K){20}
1
AA000232  CK: 1000  len: 102  ! Aa00232 Human polypeptide SEQ ID NO 141
      (R,K){20}
29: RDCFF KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
30: DCEFK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
31: CEFKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}

```

32: FFKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
33: FKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
34: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
35: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
36: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
37: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
38: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
39: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
40: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
41: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
42: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
43: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
44: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
45: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
46: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
47: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
48: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
49: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
50: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
51: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
52: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
53: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
54: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK  
55: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKKKKKKKK

AA000291 ck: 8100 len: 124 ! Aa000291 Human polypeptide SEQ ID NO 14  
(R,K){20}  
22: QHECM KKKKKKKKKKKKKKKKKKKKKKKKKKK  
AA000439 ck: 6396 len: 122 ! Aa000439 Human polypeptide SEQ ID NO 14  
(R,K){20}  
23: CLWLW KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
24: LMLVK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
25: WLVRK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
26: LVKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
27: VKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
28: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
29: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
31: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
32: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
33: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
34: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
35: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
36: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
37: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
38: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
39: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
40: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
41: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
42: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K){20}  
43: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

44: KKKKK (R,K){20}  
KKKKKKKKKKKKKKKKKKK NYPHV

AA000608 ck: 7114 len: 57 ! Aa000608 Human polypeptide SEQ ID NO 14500.

22: SLSPE (R,K){20}  
(K){20} KKKKKKKKKKKKKKKKKKK KSASS.

23: LSPEK (K){20}  
KKKKKKKKKKKKKKKKKKK SASSS

AA000619 ck: 6838 len: 117 ! Aa000619 Human polypeptide SEQ ID NO 14511.

1: (R,K){20}  
(K){20} KKKKKKKKKKKKKKKKKKK KKKKK

2: K (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKD

6: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKDG

7: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKDGG

8: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KDGCG

9: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK DGGGA

AA001368 ck: 3955 len: 123 ! Aa001368 Human polypeptide SEQ ID NO 15260.

87: ROLIT (R,K){20}  
(K){20} KKKKKKKKKKKKKKKKKKK KKKKK

88: QLITK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

89: LITKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

90: ITKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

91: TKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

92: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

93: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

94: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKD

95: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKDD

96: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKDSS

97: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KDDSG

98: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK DDSCK

AA001560 ck: 7002 len: 74 ! Aa001560 Human polypeptide SEQ ID NO 154

30: KLYHL (R,K){20}  
(K){20} KKKKKKKKKKKKKKKKKKK KKKKK

31: LYHLK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

32: YHLKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

33: HLKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKT

34: LKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKTK

35: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKTKK

36: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KTKKK

37: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK TKKKK

AA002135 ck: 8265 len: 74 ! Aa002135 Human polypeptide SEQ ID NO 160

41: MPPPP (R,K){20}  
(K){20} KKKKKKKKKKKKKKKKKKK KKKKK

42: PPPPK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

43: PPPKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

44: PPKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

45: PKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

46: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

47: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

48: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

49: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKK

50: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKK KKKKO

51: KKKK (K){20} KKKQ  
52: KKKK (K){20} KKKQ  
53: KKKK (K){20} KQ  
54: KKKK (K){20} Q  
AA002186 ck: 4844 len: 57 ! Aa002186 Human polypeptide SEQ ID NO 16078.  
(R,K){20}  
11: HCCL KKKKKKKKKKKKKKKKK KKKK  
12: CCLK (K){20} KKKK  
AA002310 ck: 704 len: 137 ! Aa002310 Human polypeptide SEQ ID NO 16202.  
(R,K){20}  
(R,K){20}  
22: HSLNL KKKKKKKKKKKKKKKR GCGVK  
AA002477 ck: 2018 len: 112 ! Aa002477 Human polypeptide SEQ ID NO 16369.  
(R,K){20}  
(K){20}  
83: LASAV KKKKKKKKKKKKKKKKK KKKK  
84: ASAVK (K){20} KKKK  
85: SAVK (K){20} KKKK  
86: AVKK (K){20} KKKK  
87: VKKK (K){20} KKKK  
88: KKKK (K){20} KKKK  
89: KKKK (K){20} KKKK  
90: KKKK (K){20} KKKK  
91: KKKK (K){20} KK  
92: KKKK (K){20} K  
93: KKKK (K){20} KKKK  
AA002733 ck: 2860 len: 132 ! Aa002733 Human polypeptide SEQ ID NO 16625.  
(R,K){20}  
(K){20}  
86: FFPSL KKKKKKKKKKKKKKKKK KKKK

87: FFSLK (K){20} KKKK  
88: FSLK (K){20} KKKK  
89: SLKK (K){20} KKKK  
90: LKKK (K){20} KKKK  
91: KKKK (K){20} KKKK  
92: KKKK (K){20} KKKK  
93: KKKK (K){20} KKKK  
94: KKKK (K){20} KKKK  
95: KKKK (K){20} KKKK  
96: KKKK (K){20} KKKR  
97: KKKK (K){20} KKKR  
98: KKKK (K){20} KKKR  
99: KKKK (K){20} KKKR  
100: KKKK (K){20} KKKR  
101: KKKK (R,K){20} KKKR  
AA002946 ck: 3649 len: 126 ! Aa002946 Human polypeptide SEQ ID NO 16  
(R,K){20}  
(K){20}  
25: DEATS KKKKKKKKKKKKKKKKK KKKK  
26: EATSK (K){20} KKKK  
27: ATSK (K){20} KKKK  
28: TSKK (K){20} KKKK  
29: SKKK (K){20} KKKK  
30: KKKK (K){20} KKKK  
31: KKKK (K){20} KKKK  
32: KKKK (K){20} KKKK

1  
33: KKKKK (K){20} KKKKK  
34: KKKKK (K){20} KKKKK  
35: KKKKK (K){20} KKKKK  
36: KKKKK (K){20} KKKKK  
37: KKKKK (K){20} KKKKK  
38: KKKKK (K){20} KKKKK  
AA002961 ck: 4320 len: 83 i Aa002961 Human polypeptide SEQ ID NO 16853  
(R,K){20}  
22: RMFSS KKKKKKKKKKKKKKKKK KTAIT  
(K){20}  
23: MFSSK KKKKKKKKKKKKKKKKK TAITK  
(K){20}  
1  
AA003006 ck: 5212 len: 102 i Aa003006 Human polypeptide SEQ ID NO 16898  
(R,K){20}  
59: HDEFP KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
60: DEFPK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
61: FFPKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
62: FPKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
63: PKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
64: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
65: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
66: KKKKK KKKKKKKKKKKKKKKKK KASSS  
(K){20}  
67: KKKKK KKKKKKKKKKKKKKKKK ASSSS  
(K){20}  
1  
AA003024 ck: 6865 len: 64 i Aa003024 Human polypeptide SEQ ID NO 16916  
(R,K){20}  
15: SFAKA KKKKKKKKKKKKKKKKK KRGGP  
(R,K){20}  
16: FAKAR KKKKKKKKKKKKKKKKK RGGPP  
(K){20}  
17: AKARK KKKKKKKKKKKKKKKKK RGPPL  
(R,K){20}

1  
AA003113 ck: 5587 len: 60 i Aa003113 Human polypeptide SEQ ID NO 170  
(R,K){20}  
33: EINML KKKKKKKKKKKKKKKKK KKKRG  
(K){20}  
34: INMLK KKKKKKKKKKKKKKKKK KKKRG  
(K){20}  
35: NMLKK KKKKKKKKKKKKKKKKK KRGGG  
(K){20}  
36: MLKKK KKKKKKKKKKKKKKKKK RGGGL  
(R,K){20}  
37: LKKKK KKKKKKKKKKKKKKKKK GGGL  
(K){20}  
1  
AA003123 ck: 6627 len: 108 i Aa003123 Human polypeptide SEQ ID NO 170  
(R,K){20}  
33: RTWRX KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
34: TWRXX KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
35: WRXKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
36: RXKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
37: XKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
38: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
39: KKKKK KKKKKKKKKKKKKKKKK KKKRG  
(K){20}  
40: KKKKK KKKKKKKKKKKKKKKKK KKKRG  
(K){20}  
41: KKKKK KKKKKKKKKKKKKKKKK KRGGG  
(K){20}  
42: KKKKK KKKKKKKKKKKKKKKKK RGGGO  
(K){20}  
43: KKKKK KKKKKKKKKKKKKKKKK GGGOX  
(R,K){20}  
1  
AA003132 ck: 3903 len: 116 i Aa003132 Human polypeptide SEQ ID NO 170  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKK KKKRG  
(K){20}  
3: KK KKKKKKKKKKKKKKKKK KKKRG  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKK KRGPP  
(K){20}



5: KKKK KKKKKKKKKKKKKKKKK RGGPP  
(R,K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKR GGPPK  
AA003152 ck: 7891 len: 35 i Aa003152 Human polypeptide SEQ ID NO 17044.  
(R,K){20}  
13: LIXYT KKKKKKKKKKKKKKKKK RGG  
(R,K){20}  
14: IXYTR KKKKKKKKKKKKKKKKKR GG  
AA003168 ck: 7233 len: 67 i Aa003168 Human polypeptide SEQ ID NO 17060.  
(R,K){20}  
17: ELLMP KKKKKKKKKKKKKKKKK KRGGA  
(K){20}  
18: LMPK KKKKKKKKKKKKKKKKK RGGAP  
(R,K){20}  
19: LMPK KKKKKKKKKKKKKKKKKR GGAPF  
AA003243 ck: 411 len: 95 i Aa003243 Human polypeptide SEQ ID NO 17135.  
(R,K){20}  
51: LHHAV KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
52: LHAV KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
53: HAVK KKKKKKKKKKKKKKKKK KKKKK  
AA003273 ck: 2981 len: 88 i Aa003273 Human polypeptide SEQ ID NO 17165.  
(R,K){20}  
43: IXYLL KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
44: XYLLK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
45: YLLK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
46: LLKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
47: LKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
50: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
51: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

52: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
53: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
54: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
55: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
56: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
57: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AA003277 ck: 2981 len: 89 i Aa003277 Human polypeptide SEQ ID NO 17  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKK KKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
AA003278 ck: 6777 len: 74 i Aa003278 Human polypeptide SEQ ID NO 17  
(R,K){20}  
47: RAVAQ KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: AVAQ KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: VAQK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
50: AQKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AA003284 ck: 9650 len: 115 i Aa003284 Human polypeptide SEQ ID NO 17  
(R,K){20}  
22: KALVS KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: ALVSK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: LVSK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: VSKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: SKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
27: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

28: KKKKK (K)(20) RGAL  
29: KKKKK (R,K)(20) GGAL  
AAO03288 ck: 1109 len: 117 ! Aao03288 Human polypeptide seq ID NO 17180.  
1  
31: SCSYP KKKKK (R,K)(20)  
32: CSYPK KKKKK (K)(20)  
33: SYPKK KKKKK (K)(20)  
34: YPKKK KKKKK (K)(20)  
35: PKKKK KKKKK (K)(20)  
36: KKKKK KKKKK (K)(20)  
37: KKKKK KKKKK (K)(20)  
38: KKKKK KKKKK (K)(20)  
39: KKKKK KKKKK (K)(20)  
40: KKKKK KKKKK (K)(20)  
41: KKKKK KKKKK (K)(20)  
42: KKKKK KKKKK (K)(20)  
43: KKKKK KKKKK (K)(20)  
44: KKKKK KKKKK (K)(20)  
45: KKKKK KKKKK (R,K)(20)  
46: KKKKK KKKKK (R,K)(20)  
47: KKKKK KKKKK (R,K)(20)  
48: KKKKK KKKKK (R,K)(20)  
49: KKKKK KKKKK (R,K)(20)  
50: KKKKK KKKKK (R,K)(20)  
51: KKKKK KKKKK (R,K)(20) KTXG

52: KKKKK (R,K)(20) KTXG  
53: KKKKK (R,K)(20) TYGG  
AAO03375 ck: 693 len: 54 ! Aao03375 Human polypeptide seq ID NO 172  
1  
25: TXLEE KKKKK (R,K)(20)  
26: XLEEK KKKKK (K)(20)  
27: LEEKK KKKKK (K)(20)  
28: EEKKK KKKKK (K)(20) APGG  
AAO03459 ck: 5217 len: 58 ! Aao03459 Human polypeptide seq ID NO 173  
1  
28: XMCVF KKKKK (R,K)(20) KDPPK  
29: MCVFK KKKKK (K)(20) DPPKK  
AAO03475 ck: 315 len: 138 ! Aao03475 Human polypeptide seq ID NO 173  
1  
44: LXLTL KKKKK (R,K)(20) KKKKK  
45: LXLTK KKKKK (K)(20) KKKKK  
46: XLTKK KKKKK (K)(20) KKKKK  
47: LTKKK KKKKK (K)(20) KKGPP  
48: TKKKK KKKKK (K)(20) KGGPR  
49: KKKKK KKKKK (K)(20) GGPRG  
AAO03615 ck: 7613 len: 128 ! Aao03615 Human polypeptide seq ID NO 175  
1  
28: PLPRX KKKKK (R,K)(20) KKTAA  
29: LPRXK KKKKK (K)(20) KTAAG  
30: PRXKK KKKKK (K)(20) TAAGG  
AAO03623 ck: 3884 len: 116 ! Aao03623 Human polypeptide seq ID NO 175  
1  
1  
(R,K)(20)  
(K)(20)

77: TPSRA KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
78: PSRAK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
79: SRAKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
80: RAKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
81: AKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
82: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
83: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
84: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
85: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
86: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
87: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
88: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

AAO03700 ck: 1279 len: 43 1 Aao03700 Human polypeptide SEQ ID NO 17592.  
(R,K){20}

19: YSQRL KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

AAO03703 ck: 7016 len: 113 1 Aao03703 Human polypeptide SEQ ID NO 17595.  
(R,K){20}

58: IMDAE KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

59: WDAEK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

60: DAEKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

61: AEKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

62: EKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

63: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

64: KKKKK KKKKKKKKKKKKKKKKKKK KSPGG  
(K){20}

65: KKKKK KKKKKKKKKKKKKKKKKKK SPGGA  
(K){20}

AAO03766 ck: 8808 len: 81 1 Aao03766 Human polypeptide SEQ ID NO 17658.  
(K){20}

1  
(R,K){20}

49: TTAC KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

50: TTACK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

51: TACK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

52: ACKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

53: CKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

54: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

55: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

56: KKKKK KKKKKKKKKKKKKKKKKKK KGGGG  
(K){20}

57: KKKKK KKKKKKKKKKKKKKKKKKK KGGGA  
(K){20}

AAO03841 ck: 8734 len: 100 1 Aao03841 Human polypeptide SEQ ID NO 17  
(R,K){20}

36: KQHYP KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

37: QHYPK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

38: HYPK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

39: YPKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

40: PKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

41: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

42: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

43: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

44: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

45: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(R,K){20}

AAO03906 ck: 4312 len: 100 1 Aao03906 Human polypeptide SEQ ID NO 17  
(R,K){20}

34: NKONQ KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

35: KONOK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

36: QNOKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
37: NOKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
38: OKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
40: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AA003967 ck: 7917 len: 53 i Aa003967 Human polypeptide SEQ ID NO 17859.  
(R,K){20}  
16: KLKKS KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: LKSKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: XKSkk KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KSKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: SKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AA004512 ck: 4277 len: 62 i Aa004512 Human polypeptide SEQ ID NO 18404.  
(R,K){20}  
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

1  
AA004619 ck: 8654 len: 39 i Aa004619 Human polypeptide SEQ ID NO 185  
(R,K){20}  
13: LITFL KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: ITFLK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: TFLKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: FLKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AA004644 ck: 2038 len: 70 i Aa004644 Human polypeptide SEQ ID NO 185  
(R,K){20}  
23: AKPPT KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}

AA004645 ck: 7825 len: 118 i Aa004645 Human polypeptide SEQ ID NO 185  
(R,K){20}  
46: IKSFL KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
47: KSFLR KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: SFLRK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: FLRRK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
50: LRKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
51: RRKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

58: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
59: KKKK KKKKKKKKKKKKKKKKK KKKR  
(K){20}  
60: KKKK KKKKKKKKKKKKKKKKK KPRG  
(K){20}  
61: KKKK KKKKKKKKKKKKKKKKK KPRG  
(K){20}  
62: KKKK KKKKKKKKKKKKKKKKK PRGG  
(K){20}

AA004647 ck: 4805 len: 58 i Aa004647 Human polypeptide SEQ ID NO 18539.  
(R,K){20}  
28: KPTRP KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
29: PTRPK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
30: TRPK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
31: RPKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
32: PKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
33: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
34: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
35: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
36: KKKK KKKKKKKKKKKKKKKKK KKS  
(K){20}  
37: KKKK KKKKKKKKKKKKKKKKK KS  
(K){20}  
38: KKKK KKKKKKKKKKKKKKKKK S  
(K){20}

AA004674 ck: 2036 len: 33 i Aa004674 Human polypeptide SEQ ID NO 18566.  
(R,K){20}  
9: WCYIT KKKKKKKKKKKKKKKKK KARG  
(K){20}  
10: CYTK KKKKKKKKKKKKKKKKK KARG  
(K){20}  
11: YITK KKKKKKKKKKKKKKKKK ARG  
(K){20}

AA004679 ck: 7190 len: 31 i Aa004679 Human polypeptide SEQ ID NO 18571.  
(R,K){20}  
8: CMFTQ KKKKKKKKKKKKKKKKK KDRG  
(K){20}  
9: MFTQK KKKKKKKKKKKKKKKKK DRG  
(K){20}

AA004682 ck: 5605 len: 60 i Aa004682 Human polypeptide SEQ ID NO 18  
(R,K){20}  
2: L KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
3: LK KKKKKKKKKKKKKKKKK KQKD  
(K){20}  
4: LKK KKKKKKKKKKKKKKKKK KQDL  
(K){20}  
5: LKK KKKKKKKKKKKKKKKKK QKDL  
(K){20}

AA004690 ck: 7157 len: 81 i Aa004690 Human polypeptide SEQ ID NO 18  
(R,K){20}  
1: KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
5: KKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
6: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
7: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
8: KKKK KKKKKKKKKKKKKKKKK PKKK  
(K){20}

AA004715 ck: 6984 len: 35 i Aa004715 Human polypeptide SEQ ID NO 18  
(R,K){20}  
8: LGSKD KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
9: GSKDK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
10: SKDK KKKKKKKKKKKKKKKKK AKKD  
(K){20}

AA004743 ck: 9412 len: 54 i Aa004743 Human polypeptide SEQ ID NO 18  
(R,K){20}  
21: KINKL KKKKKKKKKKKKKKKKK KRAA  
(K){20}  
22: INLKL KKKKKKKKKKKKKKKKK RAAAR  
(K){20}  
23: NKLKL KKKKKKKKKKKKKKKKK AAARD  
(R,K){20}

AA004747 ck: 8399 len: 39 i Aa004747 Human polypeptide SEQ ID NO 18

1  
13: RTGFV KKKKKKKKKKKKKKKKK KRRG  
(K){20}  
14: TGFVK KKKKKKKKKKKKKKKKK KRGCG  
(K){20}  
15: GFVKK KKKKKKKKKKKKKKKKK RGGGF  
(R,K){20}  
16: FVKKK KKKKKKKKKKKKKKKKK GGGF  
(R,K){20}  
AA004752 ck: 1021 len: 71 i Aa004752 Human polypeptide SEQ ID NO 18644.  
(R,K){20}  
19: OEOGL KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: EOGLK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: OGLKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: GLKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: LKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKK KKKGG  
(K){20}  
25: KKKKK KKKKKKKKKKKKKKKKK KKGCG  
(K){20}  
26: KKKKK KKKKKKKKKKKKKKKKK KGGCL  
(K){20}  
27: KKKKK KKKKKKKKKKKKKKKKK GGGLL  
(K){20}  
AA004755 ck: 5521 len: 59 i Aa004755 Human polypeptide SEQ ID NO 18647.  
(R,K){20}  
31: INSLE KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
32: NSLEK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
33: SLEKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
34: LEKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
35: EKKKK KKKKKKKKKKKKKKKKK KKKKT  
(K){20}  
36: KKKKK KKKKKKKKKKKKKKKKK KKKKT  
(K){20}  
37: KKKKK KKKKKKKKKKKKKKKKK KKT  
(K){20}  
38: KKKKK KKKKKKKKKKKKKKKKK KT  
(K){20}

1  
39: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK T  
AA004756 ck: 6952 len: 26 i Aa004756 Human polypeptide SEQ ID NO 186  
(R,K){20}  
4: FFY KKKKKKKKKKKKKKKKK SSS  
(K){20}  
AA004758 ck: 7836 len: 115 i Aa004758 Human polypeptide SEQ ID NO 186  
(R,K){20}  
7: PFYOL KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: FYOLK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: YOLKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: OLKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: LKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

27: KKKK (K){20}  
28: KKKK (K){20}  
29: KKKK (K){20}  
30: KKKK (K){20}  
31: KKKK (K){20}  
32: KKKK (K){20}  
33: KKKK (K){20}  
34: KKKK (K){20}  
35: KKKK (K){20}  
36: KKKK (K){20}  
37: KKKK (K){20}  
38: KKKK (K){20}  
39: KKKK (K){20}  
40: KKKK (K){20}  
41: KKKK (K){20}  
42: KKKK (K){20}  
43: KKKK (K){20}  
44: KKKK (K){20}  
45: KKKK (K){20}  
46: KKKK (K){20}  
47: KKKK (K){20}  
48: KKKK (K){20}  
49: KKKK (K){20}  
50: KKKK (K){20}

1 AA004764 ck: 2035 len: 54 i Aa004764 Human polypeptide SEQ ID NO 18  
(R,K){20}

9: GDSSL KKKKKKKKKKKKKKKKKKK

10: DSSLK KKKKKKKKKKKKKKKKKKK

11: SSLK KKKKKKKKKKKKKKKKKKK

12: SLKK KKKKKKKKKKKKKKKKKKK

13: LKKK KKKKKKKKKKKKKKKKKKK

14: KKKK KKKKKKKKKKKKKKKKKKK

1 AA004802 ck: 1223 len: 105 i Aa004802 Human polypeptide SEQ ID NO 18  
(R,K){20}

34: FFSRQ KKKKKKKKKKKKKKKKKKK

1

13: TPFR KKKKKKKKKKKKKKKKKKK

14: PFRK KKKKKKKKKKKKKKKKKKK

15: FRKK KKKKKKKKKKKKKKKKKKK

16: RAKK KKKKKKKKKKKKKKKKKKK

17: AKKK KKKKKKKKKKKKKKKKKKK

18: KKKK KKKKKKKKKKKKKKKKKKK

19: KKKK KKKKKKKKKKKKKKKKKKK

20: KKKK KKKKKKKKKKKKKKKKKKK

21: KKKK KKKKKKKKKKKKKKKKKKK

22: KKKK KKKKKKKKKKKKKKKKKKK

23: KKKK KKKKKKKKKKKKKKKKKKK

24: KKKK KKKKKKKKKKKKKKKKKKK

25: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

26: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
27: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
28: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
29: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
30: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
31: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
32: KKKK KKKKKKKKKKKKKKKKK KKKK  
(R,K){20}

AA004856 ck: 4861 len: 135 ! Aa004856 Human polypeptide SEQ ID NO 18748.

1  
(R,K){20}  
41: LTLTT KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
42: TLTTT KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
43: LTTTK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
44: TTKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
45: TKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
46: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
47: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
48: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
49: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
50: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
51: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
52: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
53: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
AA004872 ck: 3224 len: 93 ! Aa004872 Human polypeptide SEQ ID NO 18764.  
(R,K){20}  
8: FLYKL KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
9: LYKLK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

(K){20}  
10: YLKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
11: KLKK KKKKKKKKKKKKKKKKK KKKK  
(R,K){20}  
12: LKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

AA004873 ck: 7719 len: 66 ! Aa004873 Human polypeptide SEQ ID NO 187

1  
(R,K){20}  
21: SFLIE KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
22: FLIEK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

AA004874 ck: 5753 len: 58 ! Aa004874 Human polypeptide SEQ ID NO 187

1  
(R,K){20}  
10: SEKLP KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
11: EKLPR KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
12: KLPRK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

AA004881 ck: 8841 len: 115 ! Aa004881 Human polypeptide SEQ ID NO 187

1  
(R,K){20}  
24: ITPHP KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
25: TPHPK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
26: PHPRK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
27: HPRKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
28: PKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
29: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
30: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
31: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

AA004917 ck: 3999 len: 103 ! Aa004917 Human polypeptide SEQ ID NO 188

1  
(R,K){20}  
1: KKKKKKKKKKKKKKKKK KKKK  
(K){20}

AA004928 ck: 23 len: 43 ! Aa004928 Human polypeptide SEQ ID NO 188

1  
(R,K){20}  
19: FKKE KKKKKKKKKKKKKKKKK KKKK  
(K){20}



1

AA004969 ck: 4345 len: 57 i Aa04969 Human polypeptide SEQ ID NO 18861

(R,K){20}

(K){20}

26: LMGTS KKKKKKKKKKKKKKKKKKK

(K){20}

27: MGTSK KKKKKKKKKKKKKKKKKK

(K){20}

28: GTSK KKKKKKKKKKKKKKKKKKK

(K){20}

29: TSKK KKKKKKKKKKKKKKKKKKK

(K){20}

30: SKKK KKKKKKKKKKKKKKKKKK

(K){20}

31: KKKK KKKKKKKKKKKKKKKKKK

(K){20}

32: KKKK KKKKKKKKKKKKKKKKKK

(K){20}

33: KKKK KKKKKKKKKKKKKKKKKK

(K){20}

34: KKKK KKKKKKKKKKKKKKKKKK

(K){20}

35: KKKK KKKKKKKKKKKKKKKKK

(K){20}

36: KKKK KKKKKKKKKKKKKKKKK

(K){20}

37: KKKK KKKKKKKKKKKKKKKKK

(K){20}

38: KKKK KKKKKKKKKKKKKKKKK

AA005000 ck: 5847 len: 38 i Aa05000 Human polypeptide SEQ ID NO 18892

(R,K){20}

(K){20}

5: SGRS KKKKKKKKKKKKKKKKKKK

(K){20}

6: SGRS KKKKKKKKKKKKKKKKKK

(K){20}

7: GRSK KKKKKKKKKKKKKKKKKK

(K){20}

8: RSKK KKKKKKKKKKKKKKKKKK

(K){20}

9: SKKK KKKKKKKKKKKKKKKKKK

(K){20}

10: KKKK KKKKKKKKKKKKKKKKKK

(K){20}

11: KKKK KKKKKKKKKKKKKKKKKR

(K){20}

12: KKKK KKKKKKKKKKKKKKKKKR  
(K){20}

1

AA005081 ck: 2260 len: 40 i Aa05081 Human polypeptide SEQ ID NO 18

(R,K){20}

(K){20}

9: SRASP KKKKKKKKKKKKKKKKKK

(K){20}

10: RASPK KKKKKKKKKKKKKKKKKK

(K){20}

11: ASPKK KKKKKKKKKKKKKKKKKK

(K){20}

12: SPKK KKKKKKKKKKKKKKKKKK

(K){20}

13: PKKK KKKKKKKKKKKKKKKKKK

(K){20}

14: KKKK KKKKKKKKKKKKKKKKK

(K){20}

15: KKKK KKKKKKKKKKKKKKKSS

(K){20}

16: KKKK KKKKKKKKKKKKKKKSK

(K){20}

17: KKKK KKKKKKKKKKKKKKKSK

(K){20}

18: KKKK KKKKKKKKKKKKKKKSK

AA005130 ck: 6597 len: 116 i Aa05130 Human polypeptide SEQ ID NO 19

(R,K){20}

(K){20}

14: ETPFQ KKKKKKKKKKKKKKKKG

(K){20}

15: TPFOK KKKKKKKKKKKKKKKGG

(R,K){20}

16: PFQK KKKKKKKKKKKKKKKGFL

AA005191 ck: 4399 len: 85 i Aa05191 Human polypeptide SEQ ID NO 19

(R,K){20}

(R,K){20}

30: GYIQL KKKKKKKKKKKKKKKKGAFK

AA005260 ck: 3618 len: 55 i Aa05260 Human polypeptide SEQ ID NO 19

(R,K){20}

(K){20}

15: PHKQ KKKKKKKKKKKKKKKKKK  
(K){20}

16: HHHOK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
17: KHHOK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
18: HOKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
19: OKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKGG  
(K)[20]  
22: KKKKK KKKKKKKKKKKKKKKKKKK KGGGD  
(K)[20]  
23: KKKKK KKKKKKKKKKKKKKKKKKK GGGDF  
(K)[20]

AAO05369 ck: 1417 len: 70 ! Aao05369 Human polypeptide SEQ ID NO 19261.  
(R,K)[20]  
22: KHHET KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
23: HHETK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
24: HETKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
25: ETKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
26: TKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKRG  
(K)[20]  
28: KKKKK KKKKKKKKKKKKKKKKKKK KKRGE  
(K)[20]  
29: KKKKK KKKKKKKKKKKKKKKKKKK KRGEK  
(K)[20]  
30: KKKKK KKKKKKKKKKKKKKKKKKK RGEKK  
(R,K)[20]  
31: KKKKK KKKKKKKKKKKKKKKKKKK GEKKK  
(R,K)[20]  
AAO05372 ck: 9880 len: 43 ! Aao05372 Human polypeptide SEQ ID NO 19264.  
(R,K)[20]  
1: KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
2: K KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
3: KK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]

5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKD  
(K)[20]  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKDS  
(K)[20]  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKDSC  
(K)[20]  
16: KKKKK KKKKKKKKKKKKKKKKKKK KDSGG  
(K)[20]  
17: KKKKK KKKKKKKKKKKKKKKKKKK DSGGG  
(K)[20]

AAO05384 ck: 715 len: 23 ! Aao05384 Human polypeptide SEQ ID NO 19:  
(R,K)[20]  
4: LFS KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]

AAO05499 ck: 7813 len: 76 ! Aao05499 Human polypeptide SEQ ID NO 19  
(R,K)[20]  
7: VOPQQ KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
8: OPQQK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
9: POQKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
10: QOQKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
11: OKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)[20]

15: KKKK (K){20} KKKK  
16: KKKK (K){20} KKKK  
17: KKKK (K){20} KKKK  
18: KKKK (K){20} KKKK  
19: KKKK (K){20} KKKK  
20: KKKK (K){20} KKKK  
21: KKKK (K){20} KKKK  
22: KKKK (K){20} KKKK  
23: KKKK (K){20} KKKK  
24: KKKK (K){20} KKKK  
25: KKKK (K){20} KKKK  
AA00500 ck: 7443 len: 53 ! Aa00500 Human polypeptide SEQ ID NO 19392.  
(R,K){20}  
5: OPOS KKKKKKKKKKKKKKKKK KASQ  
(K){20}  
6: OPOS KKKKKKKKKKKKKKKKK KASQ  
(K){20}  
7: POSK KKKKKKKKKKKKKKKKK ASQK  
(K){20}  
AA00530 ck: 7354 len: 75 ! Aa00530 Human polypeptide SEQ ID NO 19422.  
(R,K){20}  
11: TTYFL KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
12: TYFLK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
13: YFLKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
14: FLKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
15: LKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

18: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
19: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
20: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
21: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
22: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
23: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
24: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
25: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
26: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
27: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
28: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
29: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
30: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
31: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
32: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
33: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
34: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
35: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
36: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
37: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
38: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
39: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
40: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
41: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
42: KKKK KKKKKKKKKKKKKKKKK ASSP  
(K){20}

1 AAO05665 ck: 432 len: 28 ! Aao05665 Human polypeptide SEQ ID NO 19557.  
(R,K){20}  
(K){20}

6: DFLQK KKKKKKKKKKKKKKKKK RGG

7: FLQK KKKKKKKKKKKKKKKKKR GG  
(R,K){20}

1 AAO06186 ck: 998 len: 88 ! Aao06186 Human polypeptide SEQ ID NO 20078.  
(R,K){20}

28: SLPP KKKKKKKKKKKKKKKKK KKKG  
(K){20}

29: LPPK KKKKKKKKKKKKKKKKK KKKG  
(K){20}

30: LPXK KKKKKKKKKKKKKKKKK KKGCG  
(K){20}

31: PXXK KKKKKKKKKKKKKKKKK KGGCG  
(K){20}

32: XXXK KKKKKKKKKKKKKKKKK CGCGF  
(K){20}

1 AAO06357 ck: 4679 len: 52 ! Aao06357 Human polypeptide SEQ ID NO 20249.  
(R,K){20}

9: FTCL KKKKKKKKKKKKKKKKK KKKIK  
(K){20}

10: TCLIK KKKKKKKKKKKKKKKKK KKIXX  
(K){20}

11: CLIK KKKKKKKKKKKKKKKKK KIKXX  
(K){20}

12: LIKK KKKKKKKKKKKKKKKKK IKXX  
(K){20}

1 AAO06429 ck: 2585 len: 71 ! Aao06429 Human polypeptide SEQ ID NO 20321.  
(R,K){20}

29: VIMX KKKKKKKKKKKKKKKKK KKKK  
(K){20}

30: IIMX KKKKKKKKKKKKKKKKK KKKG  
(K){20}

31: IMXX KKKKKKKKKKKKKKKKK KKKGR  
(K){20}

32: MXXX KKKKKKKKKKKKKKKKK KKGCG  
(K){20}

33: XXXK KKKKKKKKKKKKKKKKK KKGCG  
(K){20}

34: KKKK KKKKKKKKKKKKKKKKK GCGA  
(K){20}

1 AAO06922 ck: 7296 len: 111 ! Aao06922 Human polypeptide SEQ ID NO 20814.  
(R,K){20}  
(K){20}

13: HSEL KKKKKKKKKKKKKKKKK GCGP

1 AAO07241 ck: 3134 len: 40 ! Aao07241 Human polypeptide SEQ ID NO 211  
(R,K){20}

4: YFP KKKKKKKKKKKKKKKKK RKKX  
(K){20}

5: YFP KKKKKKKKKKKKKKKKK KKKX  
(R,K){20}

6: YFPK KKKKKKKKKKKKKKKKK KXXG  
(R,K){20}

7: FPKK KKKKKKKKKKKKKKKKK KXXG  
(R,K){20}

8: PKKK KKKKKKKKKKKKKKKKK XGCG  
(R,K){20}

1 AAO07354 ck: 6513 len: 133 ! Aao07354 Human polypeptide SEQ ID NO 212  
(R,K){20}

36: KXYHL KKKKKKKKKKKKKKKKK KKKG  
(R){20}

37: XYHLK KKKKKKKKKKKKKKKKK KGGP  
(K){20}

38: YHLKK KKKKKKKKKKKKKKKKK KGGPL  
(R){20}

39: HLKK KKKKKKKKKKKKKKKKK GGPL  
(K){20}

1 AAO07410 ck: 693 len: 80 ! Aao07410 Human polypeptide SEQ ID NO 213  
(R,K){20}

56: HKNPI KKKKKKKKKKKKKKKKK RGGF  
(K){20}

57: KNPik KKKKKKKKKKKKKKKRR CGGF  
(R,K){20}

1 AAO07412 ck: 4623 len: 166 ! Aao07412 Human polypeptide SEQ ID NO 213  
(R,K){20}

93: VNTQX KKKKKKKKKKKKKKKKK KKKK  
(K){20}

94: NTOXK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

95: TOXKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

96: OXKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

97: XXXK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

98: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

99: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

100: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
101: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
102: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
103: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
104: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
105: KKKKK KKKKKKKKKKKKKKKKKKK KKGGE  
(K){20}  
106: KKKKK KKKKKKKKKKKKKKKKKKK KGGEN  
(K){20}  
107: KKKKK KKKKKKKKKKKKKKKKKKK KGENN  
(K){20}

AA007505 ck: 3528 len: 72 ! Aa007505 Human polypeptide SEQ ID NO 21397.

(R,K){20}

14: HLVEA KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

15: LVEAK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

16: VEAKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

17: EAakk KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

18: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKK KKKDS  
(K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKK KKDSR  
(K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKK KDSRG  
(K){20}

23: KKKKK KKKKKKKKKKKKKKKKKKK DSRGG  
(K){20}

AA007509 ck: 4998 len: 140 ! Aa007509 Human polypeptide SEQ ID NO 21401.

(R,K){20}

17: KVXXE KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

18: VKXXE KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

19: KXEXE KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

20: XEXKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

20: XEXKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

21: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

30: KKKKK KKKKKKKKKKKKKKKKKKK KKGCG  
(K){20}

31: KKKKK KKKKKKKKKKKKKKKKKKK KGGGA  
(K){20}

32: KKKKK KKKKKKKKKKKKKKKKKKK GGGAS  
(K){20}

AA007594 ck: 7271 len: 93 ! Aa007594 Human polypeptide SEQ ID NO 21

(R,K){20}

9: FLLGG KKKKKKKKKKKKKKKKKKK KKKKT  
(K){20}

10: LLLGG KKKKKKKKKKKKKKKKKKK KKKTK  
(K){20}

11: LGGKK KKKKKKKKKKKKKKKKKKK KKTKK  
(K){20}

12: LGKKK KKKKKKKKKKKKKKKKKKK KTKKN  
(K){20}

13: GKKKK KKKKKKKKKKKKKKKKKKK TKKNE  
(K){20}

AA007607 ck: 2434 len: 72 ! Aa007607 Human polypeptide SEQ ID NO 21

(R,K){20}

21: KKKKE KKKKKKKKKKKKKKKKKKK GGGLF  
(K){20}

48: GLFXE KKKKKKKKKKKKKKKKKKK GGGLF  
(K){20}

AA007610 ck: 5270 len: 74 ! Aa007610 Human polypeptide SEQ ID NO 21

(R,K){20}

19: TELTI KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

(K){20}

20: ELTIK KKKKKKKKKKKKKKKKKKK  
(K){20}  
21: LTIKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
22: TIKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
23: IKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
25: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
26: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AA007755 ck: 6069 len: 90 i Aa007755 Human polypeptide SEQ ID NO 21647.

(R,K){20}  
35: KNOSS KKKKKKKKKKKKKKKKKKK  
(K){20}  
36: MOSSK KKKKKKKKKKKKKKKKKKK  
(K){20}  
37: QSSKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
38: SSKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
39: SKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
40: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
41: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
42: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AA007762 ck: 5115 len: 41 i Aa007762 Human polypeptide SEQ ID NO 21654.

(R,K){20}  
18: IPSLX KKKKKKKKKKKKKKKKKKK  
(K){20}  
19: PSLXK KKKKKKKKKKKKKKKKKKK  
(K){20}  
20: SLXKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
21: LKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
22: XKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AA007775 ck: 9827 len: 90 i Aa007775 Human polypeptide SEQ ID NO 21667.

(R,K){20}  
46: PHXPE KKKKKKKKKKKKKKKKKKK  
(K){20}

(K){20}  
47: HXPEK KKKKKKKKKKKKKKKKKKK  
(K){20}  
48: XPEKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
49: PEKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
50: EKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
51: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AA007806 ck: 1959 len: 99 i Aa007806 Human polypeptide SEQ ID NO 216

(R,K){20}  
41: KLHCI KKKKKKKKKKKKKKKKKKK  
(K){20}  
42: LHCIK KKKKKKKKKKKKKKKKKKK  
(K){20}  
43: HCIIK KKKKKKKKKKKKKKKKKKK  
(K){20}  
44: CIKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
45: IKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
46: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AA007872 ck: 8003 len: 135 i Aa007872 Human polypeptide SEQ ID NO 217

(R,K){20}  
52: TKKKQ KKKKKKKKKKKKKKKKKKK  
(K){20}  
AA007874 ck: 8007 len: 111 i Aa007874 Human polypeptide SEQ ID NO 217  
(R,K){20}  
90: EEPIS KKKKKKKKKKKKKKKKKKK  
(K){20}  
91: EPSIK KKKKKKKKKKKKKKKKKKK  
(K){20}

AA007876 ck: 8130 len: 66 i Aa007876 Human polypeptide SEQ ID NO 217

(R,K){20}  
17: IYSFE KKKKKKKKKKKKKKKKKKK  
(K){20}  
18: YSEFK KKKKKKKKKKKKKKKKKKK  
(K){20}  
19: SPEKK KKKKKKKKKKKKKKKKKKK  
(K){20}

20: FEKKK (K){20} KKKKKKKKKKKKKKKKK KKKOK  
(K){20}  
21: EKKKK KKKKKKKKKKKKKKKKKKK KKKOK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKE  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKKKK OKKEN

AAO07976 ck: 6471 len: 97 i Aao07976 Human polypeptide SEQ ID NO 21868

(R,K){20}

(K){20}

43: FPIST KKKKKKKKKKKKKKKKKKK KKKKK

44: PISTK (K){20} KKKKKKKKKKKKKKKKK KKKKK

45: ISTKK (K){20} KKKKKKKKKKKKKKKKK KKKKK

46: STKKK (K){20} KKKKKKKKKKKKKKKKK KKKKK

(K){20}

47: TKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKA

(K){20}

52: KKKKK KKKKKKKKKKKKKKKKKKK KKKAS

(K){20}

53: KKKKK KKKKKKKKKKKKKKKKKKK KKKAS

(K){20}

54: KKKKK KKKKKKKKKKKKKKKKKKK KASSG

(K){20}

55: KKKKK KKKKKKKKKKKKKKKKKKK ASSGG

AAO07981 ck: 9551 len: 81 i Aao07981 Human polypeptide SEQ ID NO 21873

(R,K){20}

(K){20}

54: KPCIY KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

55: PCIYK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

56: CIYKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

57: IYKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

58: YKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKKKK KKK  
(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
62: KKKKK KKKKKKKKKKKKKKKKKKK

AAO08018 ck: 9768 len: 48 i Aao08018 Human polypeptide SEQ ID NO 21

(R,K){20}

(R,K){20}

24: KISQV RKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

25: ISOVR KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

26: SQVRK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

27: QVRKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

28: VRKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

29: RKKKK KKKKKKKKKKKKKKKKKKK

AAO08121 ck: 5381 len: 86 i Aao08121 Human polypeptide SEQ ID NO 22

(R,K){20}

(K){20}

20: HFGLL KKKKKKKKKKKKKKKKKKK GCGAF

AAO08232 ck: 782 len: 104 i Aao08232 Human polypeptide SEQ ID NO 22

(R,K){20}

(K){20}

49: ISOQN KKKKKKKKKKKKKKKKKKK KDGGG

(K){20}

50: SQQNK KKKKKKKKKKKKKKKKKKK DGGGA

AAO08450 ck: 6228 len: 31 i Aao08450 Human polypeptide SEQ ID NO 22

(R,K){20}

(K){20}

5: NNSAK KKKKKKKKKKKKKKKKKKK KKGGA

(K){20}

6: NNSAK KKKKKKKKKKKKKKKKKKK KCGAA

(K){20}

7: NSAKK KKKKKKKKKKKKKKKKKKK GGAAP

AAO08458 ck: 1813 len: 59 i Aao08458 Human polypeptide SEQ ID NO 22

(R,K){20}

(K){20}

8: LGGWE KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
9: GGEKK KKKKKKKKKKKKKKKKKKK KKKKN  
(K)(20)  
10: GWEKK KKKKKKKKKKKKKKKKKKK KKKKN  
(K)(20)  
11: WEKKK KKKKKKKKKKKKKKKKKKK KKKKN  
(K)(20)  
12: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
13: KKKKK KKKKKKKKKKKKKKKKKKK NKKKK

AAO08460 ck: 5334 len: 62 i Aao08460 Human polypeptide SEQ ID NO 22352.

1

(R,K)(120)  
22: KEFST KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

(K)(20)  
23: EFSTK KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
24: FSTKK KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
25: STKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
26: TKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKI

(K)(20)  
31: KKKKK KKKKKKKKKKKKKKKKKKK KKKIK

(K)(20)  
32: KKKKK KKKKKKKKKKKKKKKKKKK KKKIK

(K)(20)  
33: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
34: KKKKK KKKKKKKKKKKKKKKKKKK KKKKG

AAO08466 ck: 8728 len: 107 i Aao08466 Human polypeptide SEQ ID NO 22358.

1

(R,K)(120)  
57: FMGGV KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

(K)(20)  
58: WGGV KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
59: GGVKK KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
60: GVKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
61: VKKKK KKKKKKKKKKKKKKKKKKK KRRKG

(K)(20)  
62: KKKKK KKKKKKKKKKKKKKKKKKK RRRGR

(R,K)(20)  
63: KKKKK KKKKKKKKKKKKKKKKKKK RRGKK

(R,K)(20)  
64: KKKKK KKKKKKKKKKKKKKKKKKK KGRKE

(R,K)(20)  
65: KKKKK KKKKKKKKKKKKKKKKKKK GRGEG

AAO08580 ck: 5164 len: 62 i Aao08580 Human polypeptide SEQ ID NO 224

1

(R,K)(120)  
30: PILLK KKKKKKKKKKKKKKKKKKK GGALK

AAO08591 ck: 900 len: 46 i Aao08591 Human polypeptide SEQ ID NO 224

1

(R,K)(120)  
14: LLSQ KKKKKKKKKKKKKKKKKKK GGAFK

AAO08609 ck: 5151 len: 74 i Aao08609 Human polypeptide SEQ ID NO 225

1

(R,K)(120)  
18: KCVIL KKKKKKKKKKKKKKKKKKK KGGKL

(K)(20)  
19: CVILK KKKKKKKKKKKKKKKKKKK GKRLK

AAO08612 ck: 557 len: 46 i Aao08612 Human polypeptide SEQ ID NO 225

1

(R,K)(120)  
20: TFCIM KKKKKKKKKKKKKKKKKKK KKKGG

(K)(20)  
21: FCIMK KKKKKKKKKKKKKKKKKKK KKGGA

(K)(20)  
22: CIMKK KKKKKKKKKKKKKKKKKKK KGGAL

(K)(20)  
23: IMKKK KKKKKKKKKKKKKKKKKKK GGAL

AAO08616 ck: 4555 len: 41 i Aao08616 Human polypeptide SEQ ID NO 225

1

(R,K)(120)  
6: HCALP KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
7: CALPK KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
8: ALPKK KKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)  
9: LPKKK KKKKKKKKKKKKKKKKKKK KKKKK



10: PKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKK KKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AAO08623 ck: 9243 len: 119 i Aao08623 Human polypeptide SEQ ID NO 22515.  
(R,K){20}  
(R,K){20}

21: DSKQE KKKKKKKKKKKKKKKKKKKR GGAFK  
(R,K){20}  
(R,K){20}

1  
AAO08624 ck: 5590 len: 65 i Aao08624 Human polypeptide SEQ ID NO 22516.  
(R,K){20}  
(K){20}

11: EGNMW KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

12: GNNMW KKKKKKKKKKKKKKKKKKK KKKKW  
(K){20}

13: NNMWK KKKKKKKKKKKKKKKKKKK KKKWG  
(K){20}

14: NMWKK KKKKKKKKKKKKKKKKKKK KKWGG  
(K){20}

15: WKKKK KKKKKKKKKKKKKKKKKKK KKWGA  
(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKK KKGAL  
(K){20}

AAO08631 ck: 6430 len: 135 i Aao08631 Human polypeptide SEQ ID NO 22523.  
(R,K){20}

30: TPSRA KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

31: PSRAK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

32: SRAKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

33: RAKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

34: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

40: KKKKK KKKKKKKKKKKKKKKKKKK KKKRG  
(K){20}

41: KKKKK KKKKKKKKKKKKKKKKKKK KKRGG  
(K){20}

42: KKKKK KKKKKKKKKKKKKKKKKKK KRGGG  
(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKK RCGGP  
(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKKR GGGPK  
(R,K){20}

AAO08644 ck: 3202 len: 80 i Aao08644 Human polypeptide SEQ ID NO 22:  
(R,K){20}  
(K){20}

2: E KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

3: EK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

4: EKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

5: EKKK KKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

6: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

11: KKKK (K){20} KKKK  
12: KKKK (K){20} KKKK  
13: KKKK (K){20} KKKK  
14: KKKK (K){20} KKKK  
15: KKKK (K){20} KKKK  
16: KKKK (K){20} KKKK  
17: KKKK (K){20} KKGK  
18: KKKK (K){20} KGGG  
19: KKKK (K){20} GGGG  
AA008653 ck: 6448 len: 63 i Aa008653 Human polypeptide SEQ ID NO 22545.  
(R,K){20}

12: PIRLT (K){20} KKKK  
13: IKLTJ (K){20} KKKK  
14: KLTJK (K){20} KKKK  
15: LTKKK (K){20} KKKK  
16: TKKKK (K){20} KKKK  
17: KKKK (K){20} KKGK  
18: KKKK (K){20} KKGK  
19: KKKK (K){20} KKGK  
20: KKKK (R,K){20} GAPLK  
AA008686 ck: 6936 len: 81 i Aa008686 Human polypeptide SEQ ID NO 22578.  
(R,K){20}

26: NHKK (K){20} KKKK  
27: HKKK (K){20} KKKK  
28: KKKK (K){20} KKKK  
29: KKKK (K){20} KKKK  
30: KKKK (K){20} KKKK  
31: KKKK (K){20} KKKK  
32: KKKK (K){20} KKKK  
33: KKKK (K){20} KKKK  
34: KKKK (K){20} KKEG  
35: KKKK (K){20} KKEG  
36: KKKK (K){20} KEGAL  
37: KKKK (K){20} EGALG

1  
AA008701 ck: 588 len: 46 i Aa008701 Human polypeptide SEQ ID NO 225  
(R,K){20}

7: INANS (K){20} KKKK  
8: NANSK (K){20} KKKK  
9: ANSKK (K){20} KKKK  
10: NSKKK (K){20} KKGK  
11: SKKKK (K){20} KKGK  
12: KKKK (K){20} KGGK  
13: KKKK (K){20} GGGLL

1  
AA008707 ck: 9359 len: 48 i Aa008707 Human polypeptide SEQ ID NO 225  
(R,K){20}

13: TAVTQ (K){20} KKGK  
14: AVTQK (K){20} KGGV  
15: VTQKK (K){20} GGVL

1

AA008708 ck: 9654 len: 32 i Aa008708 Human polypeptide SEQ ID NO 22600.

(R,K){20}

(K){20}

8: SMTFP KKKKKKKKKKKKKKKKKKK

(K){20}

9: MTFPK KKKKKKKKKKKKKKKKK

(K){20}

10: TFPKK KKKKKKKKKKKKKKKKK

(K){20}

11: FPKKK KKKKKKKKKKKKKKKKK

(K){20}

12: PKKKK KKKKKKKKKKKKKKKKK

(K){20}

13: KKKKK KKKKKKKKKKKKKKKKK

1

AA008717 ck: 4688 len: 99 i Aa008717 Human polypeptide SEQ ID NO 22609.

(R,K){20}

(K){20}

80: SFLLI KKKKKKKKKKKKKKKKK

1

AA008772 ck: 1431 len: 105 i Aa008772 Human polypeptide SEQ ID NO 22664.

(R,K){20}

(K){20}

23: IMNYL KKKKKKKKKKKKKKKKK

(K){20}

24: MNYLK KKKKKKKKKKKKKKKKK

(K){20}

25: NYLKK KKKKKKKKKKKKKKKKK

(K){20}

26: YLKKK KKKKKKKKKKKKKKKKK

(K){20}

27: LKKKK KKKKKKKKKKKKKKKKK

(K){20}

28: KKKKK KKKKKKKKKKKKKKKKK

(K){20}

29: KKKKK KKKKKKKKKKKKKKKKK

(K){20}

30: KKKKK KKKKKKKKKKKKKKKKK

1

AA008774 ck: 2529 len: 71 i Aa008774 Human polypeptide SEQ ID NO 22666.

(R,K){20}

(K){20}

16: RTIKL KKKKKKKKKKKKKKKKK

(K){20}

17: TIKLK KKKKKKKKKKKKKKKKK

(K){20}

18: IKLKK KKKKKKKKKKKKKKKKK

(K){20}

19: KLKKK KKKKKKKKKKKKKKKKK

1

AA008817 ck: 529 len: 28 i Aa008817 Human polypeptide SEQ ID NO 22

(R,K){20}

(K){20}

3: AX KKKKKKKKKKKKKKKKK NIWG

1

AA008820 ck: 1888 len: 84 i Aa008820 Human polypeptide SEQ ID NO 22

(R,K){20}

(K){20}

28: CMTFS KKKKKKKKKKKKKKKKK RGGF

29: MTFSK KKKKKKKKKKKKKKKKK RGGFI

1

AA008841 ck: 3529 len: 74 i Aa008841 Human polypeptide SEQ ID NO 22

(R,K){20}

(K){20}

3: PQ KKKKKKKKKKKKKKKKK KWGG

4: POK KKKKKKKKKKKKKKKKK WGGF

1

AA008857 ck: 2816 len: 49 i Aa008857 Human polypeptide SEQ ID NO 22

(R,K){20}

(K){20}

25: PPPTS KKKKKKKKKKKKKKKKK

(K){20}

26: PPTSK KKKKKKKKKKKKKKKKK

(K){20}

27: PTSKK KKKKKKKKKKKKKKKKK

(K){20}

28: TSKKK KKKKKKKKKKKKKKKKK

(K){20}

29: SKKKK KKKKKKKKKKKKKKKKK

30: KKKKK KKKKKKKKKKKKKKKKK

1

AA008913 ck: 3132 len: 52 i Aa008913 Human polypeptide SEQ ID NO 22

(R,K){20}

(K){20}

9: IIRSF KKKKKKKKKKKKKKKKK

(K){20}

10: IKSFK KKKKKKKKKKKKKKKKK

(K){20}

11: KSFKK KKKKKKKKKKKKKKKKK

12: SFKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
13: FKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
14: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
15: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
16: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
17: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
18: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
19: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
20: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
21: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
24: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
25: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
26: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
27: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
28: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
29: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
30: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
31: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
AAO08914 ck: 4641 len: 25 i Aao08914 Human polypeptide SEQ ID NO 22806.  
(R,K){20}  
(R,K){20}  
6: ILMPX KKKKKKKKKKKKKKKKKKK KKKKK  
AAO08943 ck: 5770 len: 75 i Aao08943 Human polypeptide SEQ ID NO 22835.  
(R,K){20}  
(K){20}

8: SEWAA KKKKKKKKKKKKKKKKKKK KKKKK  
9: EWAAK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
10: WAAKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
11: AAKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
AAO08994 ck: 2695 len: 29 i Aao08994 Human polypeptide SEQ ID NO 228  
(R,K){20}  
(K){20}  
9: NLPSS KKKKKKKKKKKKKKKKKKK K  
10: LPSHK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
AAO08995 ck: 6270 len: 26 i Aao08995 Human polypeptide SEQ ID NO 228  
(R,K){20}  
(K){20}  
6: MAAPP KKKKKKKKKKKKKKKKKKK I  
AAO09001 ck: 6400 len: 26 i Aao09001 Human polypeptide SEQ ID NO 228  
(R,K){20}  
(K){20}  
5: LTSS KKKKKKKKKKKKKKKKKKK KK  
6: LTSSK (K){20} KKKKKKKKKKKKKKKKKKK K  
7: TSSKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
AAO09016 ck: 4378 len: 119 i Aao09016 Human polypeptide SEQ ID NO 229  
(R,K){20}  
(R,K){20}  
10: EEEEE RRRRRRRRRRRRRRRRRR ILRKK  
AAO09057 ck: 3712 len: 58 i Aao09057 Human polypeptide SEQ ID NO 229  
(R,K){20}  
(K){20}  
5: MILN KKKKKKKKKKKKKKKKKKK KKKKK  
6: MILNK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
7: ILNKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
8: LNKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
9: NKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
10: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
AAO09066 ck: 2645 len: 29 i Aao09066 Human polypeptide SEQ ID NO 229

```
1
  (R,K){20}
4: AQQ KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
5: AQQK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
6: AQQK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
7: QQQK KKKKKKKKKKKKKKKKKKK KKK
   (K){20}
8: QKKKK KKKKKKKKKKKKKKKKKKK KK
   (K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKK K
   (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKK

AAO09072 ck: 8432 len: 42 i Aao09072 Human polypeptide SEQ ID NO 22964.
  (R,K){20}
11: LIFTL KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
12: IFLLK KKKKKKKKKKKKKKKKKKK KKKKY
   (K){20}
13: FTLKK KKKKKKKKKKKKKKKKKKK KKKYE
   (K){20}
14: TLKKK KKKKKKKKKKKKKKKKKKK KKKYE

AAO09077 ck: 9524 len: 87 i Aao09077 Human polypeptide SEQ ID NO 22969.
  (R,K){20}
26: RPLLT KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
27: FLLTK KKKKKKKKKKKKKKKKKKK KKKWK
   (K){20}
28: LLTKK KKKKKKKKKKKKKKKKKKK KKKWG
   (K){20}
29: LTKKK KKKKKKKKKKKKKKKKKKK KKKGG
   (K){20}
30: TKKKK KKKKKKKKKKKKKKKKKKK KKKGG

AAO09162 ck: 3841 len: 100 i Aao09162 Human polypeptide SEQ ID NO 23054.
  (R,K){20}
11: KTLFQ KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
12: TLFQK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
13: LFFQK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
14: FQKKK KKKKKKKKKKKKKKKKKKK KKKKK
```

```
1
15: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKRG
   (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKRG
   (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKGG
   (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKGA
   (R,K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKGA

AAO09258 ck: 7988 len: 42 i Aao09258 Human polypeptide SEQ ID NO 23
  (R,K){20}
7: LFCFP KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
8: LFCPK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
9: FCPKK KKKKKKKKKKKKKKKKKKK KKKGG
   (K){20}
10: CPKKK KKKKKKKKKKKKKKKKKKK KKKGP
   (K){20}
11: PKKKK KKKKKKKKKKKKKKKKKKK KKKGP
   (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKGP

AAO09269 ck: 7553 len: 106 i Aao09269 Human polypeptide SEQ ID NO 23
  (R,K){20}
39: NTGML KKKKKKKKKKKKKKKKKKK KKKKK
```

1 AA009457 ck: 9658 len: 32 i Aa009457 Human polypeptide SEQ ID NO 23349.  
(R,K){20}  
(K){20}  
9: ALVPO KKKKKKKKKKKKKKKKK NIKI

1 AA009819 ck: 4709 len: 80 i Aa009819 Human polypeptide SEQ ID NO 23711.  
(R,K){20}  
(K){20}  
54: CEFVX KKKKKKKKKKKKKKKKK EKGCG

1 AA010447 ck: 6342 len: 50 i Aa010447 Human polypeptide SEQ ID NO 24339.  
(R,K){20}  
(K){20}  
25: RFLKL KKKKKKKKKKKKKKKKK RKKKK  
(R,K){20}  
26: FLKLL KKKKKKKKKKKKKKKKK KKKKI  
(R,K){20}  
27: LKLKK KKKKKKKKKKKKKKKKK KKKI  
(R,K){20}  
28: KLKKK KKKKKKKKKKKKKKKKK KKI  
(R,K){20}  
29: LKKKK KKKKKKKKKKKKKKKKK KI  
(R,K){20}  
30: KKKKK KKKKKKKKKKKKKKKKK I

1 AA010451 ck: 3955 len: 82 i Aa010451 Human polypeptide SEQ ID NO 24343.  
(R,K){20}  
(K){20}  
59: SRASP KKKKKKKKKKKKKKKKK ARGG

1 AA010467 ck: 7542 len: 116 i Aa010467 Human polypeptide SEQ ID NO 24359.  
(R,K){20}  
(K){20}  
57: KCEFM KKKKKKKKKKKKKKKKK KIGGG  
(K){20}  
58: CEFMK KKKKKKKKKKKKKKKKK IGGGA

1 AA010564 ck: 9156 len: 98 i Aa010564 Human polypeptide SEQ ID NO 24456.  
(R,K){20}  
(K){20}  
18: NLLTL KKKKKKKKKKKKKKKKK KKEMP  
(K){20}  
19: LTLK KKKKKKKKKKKKKKKKK KEPPV  
(K){20}  
20: LTLKK KKKKKKKKKKKKKKKKK EMPVK

1 AA010608 ck: 9106 len: 67 i Aa010608 Human polypeptide SEQ ID NO 24500.  
(R,K){20}  
(K){20}  
42: CRLSE KKKKKKKKKKKKKKKKK KKKKD

43: RLSEK KKKKKKKKKKKKKKKKK KKKDS  
(K){20}  
44: LSEKK KKKKKKKKKKKKKKKKK KKDS  
(K){20}  
45: SEKKK KKKKKKKKKKKKKKKKK KDS  
(K){20}  
46: EKKKK KKKKKKKKKKKKKKKKK DS  
(K){20}

1 AA010638 ck: 621 len: 56 i Aa010638 Human polypeptide SEQ ID NO 245  
(R,K){20}  
(K){20}  
33: CEPQP KKKKKKKKKKKKKKKKK KARG  
(K){20}  
34: EPQPK KKKKKKKKKKKKKKKKK ARG  
(K){20}

1 AA010786 ck: 7349 len: 65 i Aa010786 Human polypeptide SEQ ID NO 246  
(R,K){20}  
(K){20}  
24: CPXFS KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: PXFSK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: XFSKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
27: FSKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
28: SKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
29: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
30: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
31: KKKKK KKKKKKKKKKKKKKKKK KKKKV  
(K){20}  
32: KKKKK KKKKKKKKKKKKKKKKK KKKKV  
(K){20}  
33: KKKKK KKKKKKKKKKKKKKKKK KKVFF  
(K){20}  
34: KKKKK KKKKKKKKKKKKKKKKK KVEFF  
(K){20}  
35: KKKKK KKKKKKKKKKKKKKKKK VPEFE  
(K){20}

1 AA010852 ck: 5147 len: 41 i Aa010852 Human polypeptide SEQ ID NO 247  
(R,K){20}  
(K){20}  
20: FSRXM KKKKKKKKKKKKKKKKK KK  
(K){20}  
21: SRXMK KKKKKKKKKKKKKKKKK K

22: RYMKK (K){20}  
KKKKKKKKKKKKKKKKKK

AAO10853 ck: 444 len: 98 i Aao10853 Human polypeptide SEQ ID NO 24745.

1 (R,K){20}

14: LRAHL (K){20}  
KKKKKKKKKKKKKKKKKK

(K){20}

15: RAHLK KKKKKKKKKKKKKKKKK

(K){20}

16: AHLK KKKKKKKKKKKKKKKKK

(K){20}

17: HLKK KKKKKKKKKKKKKKKKK

(K){20}

18: LKKK KKKKKKKKKKKKKKKKK

(K){20}

19: KKKK KKKKKKKKKKKKKKKKK

(K){20}

20: KKKK KKKKKKKKKKKKKKKKK

(K){20}

21: KKKK KKKKKKKKKKKKKKKKK

(K){20}

22: KKKK KKKKKKKKKKKKKKKKK

(K){20}

23: KKKK KKKKKKKKKKKKKKKKK

(K){20}

24: KKKK KKKKKKKKKKKKKKKKK

(K){20}

25: KKKK KKKKKKKKKKKKKKKKK

(K){20}

26: KKKK KKKKKKKKKKKKKKKKK

(K){20}

27: KKKK KKKKKKKKKKKKKKKKK

(K){20}

28: KKKK KKKKKKKKKKKKKKKKK

(R,K){20}

29: KKKK KKKKKKKKKKKKKKKKK

(R,K){20}

AAO10859 ck: 2690 len: 70 i Aao10859 Human polypeptide SEQ ID NO 24751.

(R,K){20}

44: XGITE KKKKKKKKKKKKKKKKK

(K){20}

45: GITEK KKKKKKKKKKKKKKKKK

(K){20}

AAO10933 ck: 849 len: 69 i Aao10933 Human polypeptide SEQ ID NO 24825.

(R,K){20}

33: FINTK KKKKKKKKKKKKKKKKK

(K){20}

34: INTEK KKKKKKKKKKKKKKKKK

(K){20}

35: NTEKK KKKKKKKKKKKKKKKKK

(K){20}

36: TEKKK KKKKKKKKKKKKKKKKK

(K){20}

37: EKKKK KKKKKKKKKKKKKKKKK

(K){20}

38: KKKK KKKKKKKKKKKKKKKKK

(K){20}

39: KKKK KKKKKKKKKKKKKKKKK

(K){20}

40: KKKK KKKKKKKKKKKKKKKKK

(K){20}

41: KKKK KKKKKKKKKKKKKKKKK

(K){20}

42: KKKK KKKKKKKKKKKKKKKKK

(K){20}

43: KKKK KKKKKKKKKKKKKKKKK

(K){20}

44: KKKK KKKKKKKKKKKKKKKKK

(K){20}

45: KKKK KKKKKKKKKKKKKKKKK

(R,K){20}

AAO10973 ck: 824 len: 51 i Aao10973 Human polypeptide SEQ ID NO 24

(R,K){20}

16: SYFXM KKKKKKKKKKKKKKKKK

(K){20}

17: YFYMR KKKKKKKKKKKKKKKKK

(R,K){20}

AAO10983 ck: 694 len: 51 i Aao10983 Human polypeptide SEQ ID NO 24

(K){20}

24: HPILX KKKKKKKKKKKKKKKKK

(K){20}

25: PILXK KKKKKKKKKKKKKKKKK

(K){20}

26: ILXKK KKKKKKKKKKKKKKKKK

(K){20}

27: LXXK KKKKKKKKKKKKKKKKK

(K){20}

28: XKKK KKKKKKKKKKKKKKKKK

(K){20}

29: KKKK KKKKKKKKKKKKKKKKK

(R,K){20}

AAO10995 ck: 2382 len: 122 i Aao10995 Human polypeptide SEQ ID NO 24

(K){20}

22: PCGGS KKKKKKKKKKKKKKKKK

23: CKGSK (K)(20) KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
24: KGSKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
25: GSKKK KKKKKKKKKKKKKKKKKKK KKKGG  
(K)(20)  
26: SKKKK KKKKKKKKKKKKKKKKKKK KKGCG  
(K)(20)  
27: KKKKK KKKKKKKKKKKKKKKKKKK KGGCG  
(K)(20)  
28: KKKKK KKKKKKKKKKKKKKKKKKK GCGGF

AAO10997 ck: 5315 len: 58 ! Aao10997 Human polypeptide SEQ ID NO 24889.  
(R,K)(20)  
29: LXIIQ KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
30: XIIQK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
31: IIOKK KKKKKKKKKKKKKKKKKKK KKKRG  
(K)(20)  
32: IOKKK KKKKKKKKKKKKKKKKKKK KRRGG  
(K)(20)  
33: OKKKK KKKKKKKKKKKKKKKKKKK KRGCP  
(K)(20)  
34: KKKKK KKKKKKKKKKKKKKKKKKK RCGPP  
(R,K)(20)  
35: KKKKK KKKKKKKKKKKKKKKKKKK CGPP

AAO11002 ck: 8638 len: 82 ! Aao11002 Human polypeptide SEQ ID NO 24894.  
(R,K)(20)  
5: FKRS KKKKKKKKKKKKKKKKKKK CGALL  
(R,K)(20)  
38: XFRRS KKKKKKKKKKKKKKKKKKK RGALL  
(R,K)(20)

AAO11033 ck: 3780 len: 49 ! Aao11033 Human polypeptide SEQ ID NO 24925.  
(R,K)(20)  
10: QXKFI RKKKKKKKKKKKKKKKKKK KLSKK  
(R,K)(20)  
11: XKFIR KKKKKKKKKKKKKKKKKKK LSKKI  
(K)(20)

AAO11048 ck: 7868 len: 85 ! Aao11048 Human polypeptide SEQ ID NO 24940.  
(R,K)(20)  
53: IISDP KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
54: ISDPK KKKKKKKKKKKKKKKKKKK KKKKK

55: SDPKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
56: DPKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
57: PKKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
63: KKKKK KKKKKKKKKKKKKKKKKKK KKK  
(K)(20)  
64: KKKKK KKKKKKKKKKKKKKKKKKK KK  
(K)(20)  
65: KKKKK KKKKKKKKKKKKKKKKKKK K  
(K)(20)  
66: KKKKK KKKKKKKKKKKKKKKKKKK

AAO11078 ck: 4186 len: 100 ! Aao11078 Human polypeptide SEQ ID NO 249  
(R,K)(20)  
39: PPKPD KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
40: PKPDK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
41: KPDKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
42: PDKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
43: DKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKRG  
(K)(20)  
49: KKKKK KKKKKKKKKKKKKKKKKKK KRRGG



(K){20}  
50: KKKK KKKKKKKKKKKKKKKK KRGCG  
(K){20}  
51: KKKK KKKKKKKKKKKKKKKK RGGGG  
(R,K){20}  
52: KKKK KKKKKKKKKKKKKKKR GGGGF  
AA01124 ck: 5121 len: 61 i Aa01124 Human polypeptide SEQ ID NO 25016.  
(R,K){20}  
16: KYSFL KKKKKKKKKKKKKKKK KKKK  
(K){20}  
17: VSFLK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
18: SFLK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
19: FLKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
20: LKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
21: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
22: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
23: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
24: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
25: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
26: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
27: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
28: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
29: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
30: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
31: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
32: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
33: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
34: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
35: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}  
36: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
37: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
38: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
39: KKKK KKKKKKKKKKKKKKKK KKK  
(K){20}  
40: KKKK KKKKKKKKKKKKKKKK KKK  
(K){20}  
41: KKKK KKKKKKKKKKKKKKKK K  
(K){20}  
42: KKKK KKKKKKKKKKKKKKKK  
AA01139 ck: 8807 len: 68 i Aa01139 Human polypeptide SEQ ID NO 251  
(R,K){20}  
22: VCRFP KKKKKKKKKKKKKKKK KMEGG  
(K){20}  
23: CRFPK KKKKKKKKKKKKKKKK MEGGG  
(K){20}  
AA01165 ck: 7203 len: 62 i Aa01165 Human polypeptide SEQ ID NO 251  
(R,K){20}  
18: DASMV KKKKKKKKKKKKKKKK KKKK  
(K){20}  
19: ASMVK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
20: SMVKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
21: MYKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
22: VKKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
23: KKKK KKKKKKKKKKKKKKKK KKKK  
(K){20}  
24: KKKK KKKKKKKKKKKKKKKK KASS  
(K){20}  
25: KKKK KKKKKKKKKKKKKKKK KASSS  
(K){20}  
26: KKKK KKKKKKKKKKKKKKKK ASSSQ  
AA01209 ck: 6562 len: 96 i Aa01209 Human polypeptide SEQ ID NO 251  
(R,K){20}  
63: LLLAY KKKKKKKKKKKKKKKK KKKKY  
(K){20}  
64: LLAYK KKKKKKKKKKKKKKKK KPKYL

(K){20}  
65: LAYKK KKKKKKKKKKKKKKKKKKKKK PLYLS

AA011210 ck: 863 len: 70 1 Aa011210 Human polypeptide SEQ ID NO 25102.

(R,K){20}

(K){20}

31: DLCLC KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

32: LCLCK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

33: CLCKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

34: LCKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

35: CKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

40: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

41: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

42: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

AA011214 ck: 9584 len: 68 1 Aa011214 Human polypeptide SEQ ID NO 25106.

(R,K){20}

(K){20}

19: NRVCK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

20: RVCCK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

21: VCCKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

22: CXKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

23: XKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}  
25: KKKKK KKKKKKKKKKKKKKKKKKKKK GPKLK

AA011241 ck: 1100 len: 36 1 Aa011241 Human polypeptide SEQ ID NO 251

(R,K){20}

(K){20}

8: XOTLT KKKKKKKKKKKKKKKKKKKKK KAYTS

(K){20}

9: OTLTK KKKKKKKKKKKKKKKKKKKKK AYTSK

AA011248 ck: 5013 len: 88 1 Aa011248 Human polypeptide SEQ ID NO 251

(R,K){20}

(K){20}

51: TCLYK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

52: CLYKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

53: LYKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

54: YXKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

55: XKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGAF

(K){20}

59: KKKKK KKKKKKKKKKKKKKKKKKKKK KGAFK

AA011293 ck: 6490 len: 73 1 Aa011293 Human polypeptide SEQ ID NO 251

(R,K){20}

(K){20}

46: KXDYK KKKKKKKKKKKKKKKKKKKKK KKKSG

(K){20}

47: XDYFK KKKKKKKKKKKKKKKKKKKKK KKSOG

(K){20}

48: DYFKK KKKKKKKKKKKKKKKKKKKKK KSGGG

(K){20}

49: YFKKK KKKKKKKKKKKKKKKKKKKKK SGGGA

AA011324 ck: 6024 len: 93 1 Aa011324 Human polypeptide SEQ ID NO 252

(R,K){20}

(R,K){20}

37: ISLSA KKKKKKKKKKKKKKKKKKKKK KKKGG

(K){20}

38: SLSAR KKKKKKKKKKKKKKKKKKKKK KKGCG

(K){20}

39: LSARK KKKKKKKKKKKKKKKKKKKKK KGGGG

40: SARKK (K){20}  
KKKKKKKKKKKKKKKKKK GCGGV

AA011327 ck: 3052 len: 102 i Aa011327 Human polypeptide SEQ ID NO 25219.

(R,K){20}

20: LSKEL (K){20}  
KKKKKKKKKKKKKKKKKK KKKLG

21: SKELK (K){20}  
KKKKKKKKKKKKKKKKKK KKLGD

22: KELKK (K){20}  
KKKKKKKKKKKKKKKKKK KLGDE

23: ELKKK (K){20}  
KKKKKKKKKKKKKKKKKK LGDEE

AA011342 ck: 7189 len: 106 i Aa011342 Human polypeptide SEQ ID NO 25234.

(R,K){20}

61: FFKTX (K){20}  
KKKKKKKKKKKKKKKKKK GGPLK

AA011346 ck: 7311 len: 85 i Aa011346 Human polypeptide SEQ ID NO 25238.

(R,K){20}

62: ANHWE (K){20}  
KKKKKKKKKKKKKKKKKK SGGG

AA011352 ck: 1342 len: 51 i Aa011352 Human polypeptide SEQ ID NO 25244.

(R,K){20}

25: STFKX (K){20}  
KKKKKKKKKKKKKKKKKK KNPPS

26: IFKKX (K){20}  
KKKKKKKKKKKKKKKKKK NPPSL

AA011705 ck: 7660 len: 101 i Aa011705 Human polypeptide SEQ ID NO 25597.

(R,K){20}

17: WAPLQ (K){20}  
KKKKKKKKKKKKKKKKKK KKKGG

18: APLQK (K){20}  
KKKKKKKKKKKKKKKKKK KKGGA

19: PLQKK (K){20}  
KKKKKKKKKKKKKKKKKK KGGAP

20: LQKKK (K){20}  
KKKKKKKKKKKKKKKKKK GGAPL

AA011799 ck: 8286 len: 39 i Aa011799 Human polypeptide SEQ ID NO 25691.

(R,K){20}

7: NPVVS (R,K){20}  
RKKKKKKKKKKKKKKKKKK KKKKK

8: PPVSR (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

9: PVSRR (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

10: VSRKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

11: SRKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

12: RKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

13: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKGG

14: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKGCG

15: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KGGGL

16: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK GGGL

AA011820 ck: 8725 len: 42 i Aa011820 Human polypeptide SEQ ID NO 25

(R,R){20}

17: NFHAL (K){20}  
KKKKKKKKKKKKKKKKKK SSSLR

AA011828 ck: 4807 len: 41 i Aa011828 Human polypeptide SEQ ID NO 25

(R,R){20}

1: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

2: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

3: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

4: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKRR

5: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKRG

6: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKRGG

7: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KRGGG

8: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK RGGGF

9: KKKKK (R,K){20}  
KKKKKKKKKKKKKKKKKK RGGFK

AA011831 ck: 1761 len: 29 i Aa011831 Human polypeptide SEQ ID NO 25

(R,K){20}

1: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKA

2: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKAG

3: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKGAG

(K){20}



```
32: SNSSL KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
33: NSSLK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
34: SSLKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
35: SLKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
36: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKGG
      (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKK KKGCG
      (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKK KKGCG
      (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKK GGGGV
      (K){20}

AA011888 ck: 6488 len: 31 i Aa011888 Human polypeptide SEQ ID NO 25780
      (R,K){20}
2: D KKKKKKKKKKKKKKKKKKKKK KAGCG
      (K){20}
3: DK KKKKKKKKKKKKKKKKKKKKK AGGGL
      (K){20}

AA011903 ck: 8876 len: 48 i Aa011903 Human polypeptide SEQ ID NO 25795
      (R,K){20}
17: VIKYV KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
18: IKYVK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
19: KYVKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
20: YVKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
21: VKKKK KKKKKKKKKKKKKKKKKKK KKKKK
```

```
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KKK
      (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKK KK
      (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKK K
      (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKK
      (K){20}

AA011999 ck: 3954 len: 47 i Aa011999 Human polypeptide SEQ ID NO 25
      (R,K){20}
17: IDEGL KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
18: DEGLK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
19: EGLKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
20: GLKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
21: LKKKK KKKKKKKKKKKKKKKKKKK KKKGG
      (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKGGA
      (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKGAP
      (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KGAP
      (K){20}

AA012098 ck: 8434 len: 39 i Aa012098 Human polypeptide SEQ ID NO 25
      (R,K){20}
13: TCKFG KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
14: CKFGK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
15: KFGKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
16: FGKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
17: GKKKK KKKKKKKKKKKKKKKKKKK KKK
      (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KK
```

19: KKKKK (K)(20) KKKKKKKKKKKKKKKKK K  
20: KKKKK (K)(20) KKKKKKKKKKKKKKKKK

AAO12105 ck: 6255 len: 80 i Aao12105 Human polypeptide SEQ ID NO 25997

1

(R,K)(20)  
(K)(20)

53: EATSL KKKKKKKKKKKKKKKKK KKKK

54: ATSLK KKKKKKKKKKKKKKKKK KKKKA  
(K)(20)

55: TSLKK KKKKKKKKKKKKKKKKK KKKAR  
(K)(20)

56: SLKKK KKKKKKKKKKKKKKKKK KKARG  
(K)(20)

57: LKKKK KKKKKKKKKKKKKKKKK KKARG  
(K)(20)

58: KKKKK (K)(20) KKKKKKKKKKKKKKKKK ARG

AAO12179 ck: 1295 len: 69 i Aao12179 Human polypeptide SEQ ID NO 26071

1

(R,K)(20)  
(K)(20)

35: PKEFSV KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

36: KFSVK KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

37: FSVKK KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

38: SVKKK KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

39: VAKKK KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

40: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

41: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

42: KKKKK KKKKKKKKKKKKKKKKK KKKKG  
(K)(20)

43: KKKKK KKKKKKKKKKKKKKKKK KKKGG  
(K)(20)

44: KKKKK KKKKKKKKKKKKKKKKK KKKGG  
(K)(20)

45: KKKKK KKKKKKKKKKKKKKKKK KGGGA  
(K)(20)

46: KKKKK KKKKKKKKKKKKKKKKK GGGA  
(K)(20)

AAO12180 ck: 9840 len: 67 i Aao12180 Human polypeptide SEQ ID NO 26072

1

(R,K)(20)

1

38: CSYLP (K)(20) KKKKKKKKKKKKKKKKK KKKKK

39: SYLPK (K)(20) KKKKKKKKKKKKKKKKK KKKKK

40: YLPKK (K)(20) KKKKKKKKKKKKKKKKK KKKKK

41: LPKKK (K)(20) KKKKKKKKKKKKKKKKK KKKKK

42: PKKKK (K)(20) KKKKKKKKKKKKKKKKK KKKKK

43: KKKKK (K)(20) KKKKKKKKKKKKKKKKK KKKKK

44: KKKKK (K)(20) KKKKKKKKKKKKKKKKK KKKK

45: KKKKK (K)(20) KKKKKKKKKKKKKKKKK KKK

46: KKKKK (K)(20) KKKKKKKKKKKKKKKKK KK

47: KKKKK (K)(20) KKKKKKKKKKKKKKKKK K

48: KKKKK (K)(20) KKKKKKKKKKKKKKKKK

AAO12187 ck: 4700 len: 60 i Aao12187 Human polypeptide SEQ ID NO 260

30: SCKKL (R,K)(20) KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

31: CKKLK (K)(20) KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

32: KKLKK (K)(20) KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

33: XLKKK (K)(20) KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

34: LKKKK (K)(20) KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

35: KKKKK (K)(20) KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

36: KKKKK (K)(20) KKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

37: KKKKK (K)(20) KKKKKKKKKKKKKKKKK KKKK  
(K)(20)

38: KKKKK (K)(20) KKKKKKKKKKKKKKKKK KKK  
(K)(20)

39: KKKKK (K)(20) KKKKKKKKKKKKKKKKK KK  
(K)(20)

40: KKKKK (K)(20) KKKKKKKKKKKKKKKKK K  
(K)(20)

41: KKKKK (K)(20) KKKKKKKKKKKKKKKKK

AA012203 ck: 4083 len: 41 i Aa012203 Human polypeptide SEQ ID NO 26095.

(R,K){20}

6: SMCC L KKKKKKKKKKKKKKKKKKK

(K){20}

7: WCCL KKKKKKKKKKKKKKKKKKK

(K){20}

8: CCL KKKKKKKKKKKKKKKKKKK

(K){20}

9: CL KKKKKKKKKKKKKKKKKKK

AA012215 ck: 6903 len: 73 i Aa012215 Human polypeptide SEQ ID NO 26107.

(R,K){20}

50: PPLP KKKKKKKKKKKKKKKKKKK

(K){20}

51: PPLP KKKKKKKKKKKKKKKKKKK

AA012243 ck: 8474 len: 39 i Aa012243 Human polypeptide SEQ ID NO 26135.

(R,K){20}

11: MISFI KKKKKKKKKKKKKKKKKKK

(K){20}

12: ISFI KKKKKKKKKKKKKKKKKKK

(K){20}

13: SFIK KKKKKKKKKKKKKKKKKKK

(K){20}

14: FI KKKKKKKKKKKKKKKKKKK

(K){20}

15: IK KKKKKKKKKKKKKKKKKKK

(K){20}

16: KKKKKKKKKKKKKKKKKKK GPL

AA012250 ck: 1538 len: 69 i Aa012250 Human polypeptide SEQ ID NO 26142.

(R,K){20}

31: SPSNL KKKKKKKKKKKKKKKKKKK

(K){20}

32: PSNL KKKKKKKKKKKKKKKKKKK

(K){20}

33: SNL KKKKKKKKKKKKKKKKKKK

(K){20}

34: NL KKKKKKKKKKKKKKKKKKK

(K){20}

35: L KKKKKKKKKKKKKKKKKKK

(K){20}

36: KKKKKKKKKKKKKKKKKKK

(K){20}

37: KKKKKKKKKKKKKKKKKKK

38: KKKKKKKKKKKKKKKKKKK

(K){20}

39: KKKKKKKKKKKKKKKKKKK

(K){20}

40: KKKKKKKKKKKKKKKKKKK

(K){20}

41: KKKKKKKKKKKKKKKKKKK

(K){20}

42: KKKKKKKKKKKKKKKKKKK

(K){20}

43: KKKKKKKKKKKKKKKKKKK

(K){20}

44: KKKKKKKKKKKKKKKKKKK

(R,K){20}

45: KKKKKKKKKKKKKKKKKKK

AA012274 ck: 2739 len: 37 i Aa012274 Human polypeptide SEQ ID NO 26

(R,K){20}

14: KGLIN KKKKKKKKKKKKKKKKKKK

(K){20}

15: GLINK KKKKKKKKKKKKKKKKKKK

(K){20}

16: LINK KKKKKKKKKKKKKKKKKKK

(K){20}

17: INKK KKKKKKKKKKKKKKKKKKK

AA012280 ck: 7481 len: 66 i Aa012280 Human polypeptide SEQ ID NO 26

(R,K){20}

29: HPLP KKKKKKKKKKKKKKKKKKK

(K){20}

30: IPPL KKKKKKKKKKKKKKKKKKK

(K){20}

31: PPLK KKKKKKKKKKKKKKKKKKK

(K){20}

32: PLK KKKKKKKKKKKKKKKKKKK

(K){20}

33: LKK KKKKKKKKKKKKKKKKKKK

(K){20}

34: KKKKKKKKKKKKKKKKKKK

(K){20}

35: KKKKKKKKKKKKKKKKKKK

(K){20}

36: KKKKKKKKKKKKKKKKKKK

(K){20}

37: KKKKKKKKKKKKKKKKKKK

(K){20}

38: KKKKKKKKKKKKKKKKKKK

(K){20}  
39: KKKKK KKKKKKKKKKKKKKKKK KKKSG  
(K){20}  
40: KKKKK KKKKKKKKKKKKKKKKK KKSOG  
(K){20}  
41: KKKKK KKKKKKKKKKKKKKKKK KSGGG  
(K){20}  
42: KKKKK KKKKKKKKKKKKKKKKK SGGGA  
AAO12447 ck: 5017 len: 47 i Aao12447 Human polypeptide SEQ ID NO 26339.  
(R,K){20}  
14: HLVCE KKKKKKKKKKKKKKKKK KRGGA  
(K){20}  
15: LVCEK KKKKKKKKKKKKKKKKK RGGAL  
(K){20}  
16: VCEKK KKKKKKKKKKKKKKKKK RGALK  
(R,K){20}  
AAO12476 ck: 837 len: 104 i Aao12476 Human polypeptide SEQ ID NO 26368.  
(R,K){20}  
65: RECHQ KKKKKKKKKKKKKKKKK KGFEF  
(K){20}  
66: FCHQK KKKKKKKKKKKKKKKKK GFEFF  
(K){20}  
AAO12548 ck: 6973 len: 74 i Aao12548 Human polypeptide SEQ ID NO 26440.  
(R,K){20}  
38: AVLPL KKKKKKKKKKKKKKKKK KKGCG  
(K){20}  
39: VLPLK KKKKKKKKKKKKKKKKK KGGGV  
(K){20}  
40: LPLKK KKKKKKKKKKKKKKKKK GGGVF  
(K){20}  
AAO12553 ck: 1903 len: 33 i Aao12553 Human polypeptide SEQ ID NO 26445.  
(R,K){20}  
9: SCCFI KKKKKKKKKKKKKKKKK KRGAP  
(K){20}  
10: CCFIK KKKKKKKKKKKKKKKKK RGAP  
(R,K){20}  
11: CFIKK KKKKKKKKKKKKKKKKK GAP  
(R,K){20}  
AAO13164 ck: 2798 len: 71 i Aao13164 Human polypeptide SEQ ID NO 27056.  
(R,K){20}  
35: RPPLX RRRRRRRRRRRRRRRRRR KEMFK  
(R,K){20}  
36: PPLXR KKKKKKKKKKKKKKKKK EMFKR

1  
AAO13576 ck: 4846 len: 99 i Aao13576 Human polypeptide SEQ ID NO 274  
(R,K){20}  
53: TNNLI KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
54: NNLIK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
55: NLIKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
56: LIKKK KKKKKKKKKKKKKKKKK KKKKG  
(K){20}  
57: IKKKK KKKKKKKKKKKKKKKKK KKKGG  
(K){20}  
58: KKKKK KKKKKKKKKKKKKKKKK KKGCG  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKK KGGGP  
(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKK GGCGP  
(K){20}  
AAO13785 ck: 6241 len: 100 i Aao13785 Human polypeptide SEQ ID NO 276  
(R,K){20}  
41: LYAPP KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
42: YAPPK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
43: APPKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
44: PPKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
45: PKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
46: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
50: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
51: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKK KKKKA  
(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKK KKKAG



(K) [20]  
55: KKKKK KKKKKKKKKKKKKKKKK KACGC  
(K) [20]  
56: KKKKK KKKKKKKKKKKKKKKKK KACGC  
(K) [20]  
57: KKKKK KKKKKKKKKKKKKKKKK ACGGC  
  
AAU17983 ck: 7016 len: 315 ! Aau17983 Human Immunoglobulin polypeptide S  
(R,K) [20]  
273: QVFAP KKKKKKKKKKKKKKKKK KGCGRS  
(R,K) [20]  
274: VFAPR KKKKKKKKKKKKKKKKK GGRSR  
(K) [20]  
  
AAU18049 ck: 7611 len: 315 ! Aau18049 Human Immunoglobulin polypeptide S  
(R,K) [20]  
273: QVFAP KKKKKKKKKKKKKKKKK KGCGRS  
(R,K) [20]  
274: VFAPR KKKKKKKKKKKKKKKKK GGRSR  
(K) [20]  
  
AAM14119 ck: 5383 len: 86 ! Aam14119 Peptide #553 encoded by probe for  
(R,K) [20]  
15: RRRRG RRRRRKKKKKKKKKKKK KRRRR  
(R,K) [20]  
16: RRRGR RRRRRKKKKKKKKKKKK KRRRR  
(R,K) [20]  
17: RRGRR RRRRRKKKKKKKKKKKK KRRRR  
(R,K) [20]  
18: RGRRR RRRRRKKKKKKKKKKKK KRRRR  
(R,K) [20]  
19: GRRRR RRRRRKKKKKKKKKKKK KRRRR  
(R,K) [20]  
20: RRRRR RRRRRKKKKKKKKKKKK KRRRR  
(R,K) [20]  
21: RRRRK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
22: RRRKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
23: RRRKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
24: RKKKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
25: KKKKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
26: KKKKK RRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
27: KKKKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]

28: KKKKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
29: KKKKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
30: KKKKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
31: KKKKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
32: KKKKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
33: KKKKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
34: KKKKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
35: KKKKK KRRRRKKKKKKKKKKKK RRRRR  
(R,K) [20]  
36: KKKKK KRRRRKKKKKKKKKKKK RRRRR  
(R) [20]  
37: KKKKK RRRRRKKKKKKKKKKKK RRRRR  
(R) [20]  
38: KKKKK RRRRRKKKKKKKKKKKK RRRRR  
(R) [20]  
39: KKKKK RRRRRKKKKKKKKKKKK RRRRR  
(R) [20]  
40: KKKKK RRRRRKKKKKKKKKKKK RRRRR  
(R) [20]  
41: KRRRR RRRRRKKKKKKKKKKKK RRRRR  
(R) [20]  
42: RRRRR RRRRRKKKKKKKKKKKK RRRRR  
(R) [20]  
43: RRRRR RRRRRKKKKKKKKKKKK RRRRR  
(R) [20]  
44: RRRRR RRRRRKKKKKKKKKKKK RRRRR  
(R) [20]  
45: RRRRR RRRRRKKKKKKKKKKKK RRRRR  
  
AAM14961 ck: 1334 len: 86 ! Aam14961 Peptide #1395 encoded by probe  
(R,K) [20]  
57: EEEEG RRRRRKKKKKKKKKKKK KKKKK  
(R,K) [20]  
58: EEEGR RRRRRKKKKKKKKKKKK KKKKK  
(R,K) [20]  
59: EEEGR RRRRRKKKKKKKKKKKK KKKKK  
(R,K) [20]  
60: EGRRR KRRRRKKKKKKKKKKKK KKKKK  
(R,K) [20]  
61: GRRRK KRRRRKKKKKKKKKKKK KKKKK  
(R,K) [20]

62: RRRKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K)(20)  
63: RKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(R,K)(20)  
64: RKKKK RKKKKKKKKKKKKKKKKKK KKK  
(R,K)(20)  
65: KKKKK KKKKKKKKKKKKKKKKKKK KK  
(R,K)(20)  
66: KKKKK RKKKKKKKKKKKKKKKKKK K  
(K)(20)  
67: KKKRR KKKKKKKKKKKKKKKKKKK

AAM15038 ck: 9082 len: 167 ! Aam15038 Peptide #1472 encoded by probe for  
(R,K)(20)  
33: EGRGG RRRRRRRRRRRRRRRRRRR RRGGG  
(R,K)(20)  
34: EGRGR RRRRRRRRRRRRRRRRRRR RGGGR  
(R,K)(20)  
35: GGGRR RRRRRRRRRRRRRRRRRRR GGGRR  
(R,K)(20)

AAM15527 ck: 2276 len: 89 ! Aam15527 Peptide #1961 encoded by probe for  
(R,K)(20)  
23: EEEEE KKKKKKKKKKKKKKKKKKK EEEKK  
(R,K)(20)  
46: KKEEE KKKKKKKKKKKKKKKKKKK KRKEE  
(K)(20)  
47: KEEKK KKKKKKKKKKKKKKKKKKK RKEEE  
(R,K)(20)  
48: EEEKK KKKKKKKKKKKKKKKKKKK KEEEE  
(R,K)(20)  
49: EEEKK KKKKKKKKKKKKKKKKKKK EEEEE  
(R,K)(20)

AAM15826 ck: 1939 len: 130 ! Aam15826 Peptide #2260 encoded by probe for  
(R,K)(20)  
42: EGRKE RRRRRRRRRRRRRRRRRRR RRRRR  
(R,K)(20)  
43: GKKER RRRRRRRRRRRRRRRRRRR RRRRR  
(R,K)(20)  
44: RKKER RRRRRRRRRRRRRRRRRRR RRRRR  
(R,K)(20)  
45: KERRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R,K)(20)  
46: ERRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R,K)(20)  
47: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R,K)(20)  
48: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R,K)(20)

49: RRRRR RRRRRRRRRRRRRRRRRRR KKKKK  
(R,K)(20)  
50: RRRRR RRRRRRRRRRRRRRRRRRR KKKKE  
(R,K)(20)  
51: RRRRR RRRRRRRRRRRRRRRRRRR KKKKE  
(R,K)(20)  
52: RRRRR RRRRRRRRRRRRRRRRRRR KKEEE  
(R,K)(20)  
53: RRRRR RRRRRRRRRRRRRRRRRRR KEEEE  
(R,K)(20)  
54: RRRRR RRRRRRRRRRRRRRRRRRR EEEEE  
(R,K)(20)

AAM16123 ck: 3607 len: 88 ! Aam16123 Peptide #2557 encoded by probe for  
(R,K)(20)  
39: ERKRE KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
40: RKREK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
41: KREKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
42: REKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
43: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)

56: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 57: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 58: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 59: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 60: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 61: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 62: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 63: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 64: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 65: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 66: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKK  
 67: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KK  
 68: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK K  
 69: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK

AAM16691 ck: 3937 len: 85 ! Aam16691 Peptide #3125 encoded by probe for

1: (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 2: (K) [20]  
 K KKKKKKKKKKKKKKKKKKKKK KKKK  
 3: (K) [20]  
 KK KKKKKKKKKKKKKKKKKKKKK KKKK  
 4: (K) [20]  
 KKK KKKKKKKKKKKKKKKKKKKKK KKKK  
 5: (K) [20]  
 KKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
 6: (K) [20]  
 KKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
 7: (K) [20]  
 KKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
 8: (K) [20]  
 KKKK KKKKKKKKKKKKKKKKKKKKK KKKK  
 9: (K) [20]  
 KKKK KKKKKKKKKKKKKKKKKKKKK KKKK

10: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKEEE  
 11: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK EEEEX

AAM17388 ck: 2686 len: 71 ! Aam17388 Peptide #3822 encoded by probe

20: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKR  
 21: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKR  
 22: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKR  
 23: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKR  
 24: KKKK (K) [20]  
 KKKKKKKKKKKKKKKKKKKKK RKKR  
 25: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 26: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK RKKK  
 27: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 28: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 29: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 30: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 31: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 32: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 33: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 34: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 35: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 36: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 37: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 38: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 39: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK  
 40: KKKK (R,K) [20]  
 KKKKKKKKKKKKKKKKKKKKK KKKK

1

```

41: KKKKK (R,K)(20) KKKKK
42: KKKKK (R,K)(20) KKKKK
43: KKKKK (R,K)(20) KKKKK
44: KKKKK (R,K)(20) KKKKK
45: KKKKK (R,K)(20) KKKKK
46: KKKKK (R,K)(20) KKKSA
47: KKKKK (K)(20) KKSAA
48: KKKKK (K)(20) KSAH
49: KKKKK (K)(20) SAH

AAM18532 ck: 1560 len: 88 ! Aam18532 Peptide #4966 encoded by probe for
(R,K)(20)
43: KKKKK (R,K)(20) KKKKK
44: KKKKK (R,K)(20) KKKKK
45: KKKKK (R,K)(20) KKKKK
46: KKKKK (R,K)(20) KKKKK
47: KKKKK (R,K)(20) KKKKK
48: KKKKK (R,K)(20) KKKKK
49: KKKKK (R,K)(20) KKKKK
50: KKKKK (R,K)(20) KKKKK
51: KKKKK (R,K)(20) KKKKK
52: KKKKK (R,K)(20) KKKKK
53: KKKKK (R,K)(20) KKKKK
54: KKKKK (R,K)(20) KKKKK
55: KKKKK (R,K)(20) KKKKK
56: KKKKK (R,K)(20) KKKKK

```

1

```

57: KKKKK (R,K)(20) KKKKK
58: KKKKK (R,K)(20) KKKKK
59: KKKKK (R,K)(20) KKKKK
60: KKKKK (R,K)(20) KKKKK
61: KKKKK (R,K)(20) KKKKK
62: KKKKK (R,K)(20) KKKKK

```

1

```

AAM20089 ck: 8343 len: 66 ! Aam20089 Peptide #5523 encoded by probe
(R,K)(20)
6: KKKKK (R,K)(20) KKKKK
7: KKKKK (R,K)(20) KKKKK
8: KKKKK (R,K)(20) KKKKK
9: KKKKK (R,K)(20) KKKKK

AAM21941 ck: 4895 len: 51 ! Aam21941 Peptide #8375 encoded by probe
(R,K)(20)
18: KKKKK (R,K)(20) KKKKK
19: KKKKK (R,K)(20) KKKKK
20: KKKKK (R,K)(20) KKKKK

AAM26526 ck: 5383 len: 86 ! Aam26526 Peptide #563 encoded by probe
(R,K)(20)
15: KKKKK (R,K)(20) KKKKK
16: KKKKK (R,K)(20) KKKKK
17: KKKKK (R,K)(20) KKKKK
18: KKKKK (R,K)(20) KKKKK
19: KKKKK (R,K)(20) KKKKK
20: KKKKK (R,K)(20) KKKKK
21: KKKKK (R,K)(20) KKKKK

```

22: RRRKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
23: RRRKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
24: RKKKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
25: KKKKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
26: KKKKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
27: KKKKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
28: KKKRK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
29: KKKKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
30: KKKKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
31: RKKKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
32: KKKKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
33: KKKKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
34: KKKRK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
35: KKKKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
36: KKKKK (R,K)(20) KKKKKKKKKKKKKRRRRR RRRRR  
37: RKKKK (R)(20) RRRRRRRRRRRRRRRRRR RRRRR  
38: KKKKK (R)(20) RRRRRRRRRRRRRRRRRR RRRRR  
39: KKKRR (R)(20) RRRRRRRRRRRRRRRRRR RRRRR  
40: KKKRR (R)(20) RRRRRRRRRRRRRRRRRR RRRRR  
41: KRRRR (R)(20) RRRRRRRRRRRRRRRRRR RRRRR  
42: RRRRR (R)(20) RRRRRRRRRRRRRRRRRR RRRRR  
43: RRRRR (R)(20) RRRRRRRRRRRRRRRRRR RRRRR  
44: RRRRR (R)(20) RRRRRRRRRRRRRRRRRR RRRRR  
45: RRRRR (R)(20) RRRRRRRRRRRRRRRRRR RRRRR

1

AAM27395 ck: 1334 len: 86 ! Aam27395 Peptide #1432 encoded by probe

(R,K)(20)  
(R,K)(20)

57: EEEEG RRRKKKKKKKKKKKKKKKKKK KKKKK

(R,K)(20)

58: EEEGR RRRKKKKKKKKKKKKKKKKKK KKKKK

(R,K)(20)

59: EEGRR RRRKKKKKKKKKKKKKKKKKK KKKKK

(R,K)(20)

60: EGRRR KKKKKKKKKKKKKKKKKKK KKKKK

(R,K)(20)

61: GRRRK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K)(20)

62: RRRKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K)(20)

63: RRRKK RRRKKKKKKKKKKKKKKKKKK KKKKK

(R,K)(20)

64: RKKKK RRRKKKKKKKKKKKKKKKKKK KKKKK

(R,K)(20)

65: KKKKK RRRKKKKKKKKKKKKKKKKKK KKKKK

(R,K)(20)

66: KKKRK RRRKKKKKKKKKKKKKKKKKK KKKKK

(K)(20)

67: KRRKR KRRKKKKKKKKKKKKKKKKKK KKKKK

1

AAM27479 ck: 9082 len: 167 ! Aam27479 Peptide #1516 encoded by probe

(R,K)(20)  
(R,K)(20)

33: EEGRG RRRRRRRRRRRRRRRRRR RRGCG

(R,K)(20)

34: EGRGR RRRRRRRRRRRRRRRRRR RGGGR

(R,K)(20)

35: GGRGR RRRRRRRRRRRRRRRRRR GGGRR

1

AAM28016 ck: 2276 len: 89 ! Aam28016 Peptide #2053 encoded by probe

(R,K)(20)  
(R,K)(20)

23: EEEEE KKKKKKKKKKKKKKKKKKK EEEKK

(K)(20)

46: KEEEE KKKKKKKKKKKKKKKKKKK KKEEE

(K)(20)

47: KEEKK KKKKKKKKKKKKKKKKKKK RKEEE

(R,K)(20)

48: EEEKK KKKKKKKKKKKKKKKKKKK KEEEE

(R,K)(20)

49: EEEKK KKKKKKKKKKKKKKKKKKK EEEEE

1

AAM28336 ck: 1939 len: 130 ! Aam28336 Peptide #2373 encoded by probe

(R,K)(20)  
(R,K)(20)

42: EGAKK RRRRRRRRRRRRRRRRRR  
(R,K){20}  
43: GKRR RRRRRRRRRRRRRRRRRR  
(R,K){20}  
44: KKRR RRRRRRRRRRRRRRRRRR  
(R,K){20}  
45: KERR RRRRRRRRRRRRRRRRRR  
(R,K){20}  
46: ERRR RRRRRRRRRRRRRRRRRR  
(R,K){20}  
47: RRRR RRRRRRRRRRRRRRRRRR  
(R,K){20}  
48: RRRR RRRRRRRRRRRRRRRRRR  
(R,K){20}  
49: RRRR RRRRRRRRRRRRRRRRRR  
(R,K){20}  
50: RRRR RRRRRRRRRRRRRRRRRR  
(R,K){20}  
51: RRRR RRRRRRRRRRRRRRRRRR  
(R,K){20}  
52: RRRR RRRRRRRRRRRRRRRRRR  
(R,K){20}  
53: RRRR RRRRRRRRRRRRRRRRRR  
(R,K){20}  
54: RRRR RRRRRRRRRRRRRRRRRR  
(R,K){20}

AAM28616 ck: 3607 len: 88 i Aam28616 Peptide #2653 encoded by probe for  
(R,K){20}  
39: ERRE KKKKKKKKKKKKKKKKKK  
(K){20}  
40: RREK KKKKKKKKKKKKKKKKKK  
(K){20}  
41: KREK KKKKKKKKKKKKKKKKKK  
(K){20}  
42: REKK KKKKKKKKKKKKKKKKKK  
(K){20}  
43: EKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
44: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
45: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
46: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
47: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
48: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}

49: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
50: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
51: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
52: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
53: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
54: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
55: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
56: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
57: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
58: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
59: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
60: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
61: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
62: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
63: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
64: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
65: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
66: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
67: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
68: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}  
69: KKKK KKKKKKKKKKKKKKKKKK  
(K){20}

AAM29178 ck: 3937 len: 85 i Aam29178 Peptide #3215 encoded by probe  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKKK  
(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K)(20)  
7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKE  
(K)(20)  
8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKE  
(K)(20)  
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKEE  
(K)(20)  
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KEEEE  
(K)(20)  
11: KKKKK KKKKKKKKKKKKKKKKKKKKK EEEEX  
(K)(20)  
AAM30989 ck: 1560 len: 88 ! Aam30989 Peptide #5026 encoded by probe for  
(R,K)(20)  
43: RERGG RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
44: ERGGR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
45: RRGGR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
46: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
47: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)

57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRT  
(R)(20)  
60: RRRRR RRRRRRRRRRRRRRRRRRRR RRRTN  
(R)(20)  
61: RRRRR RRRRRRRRRRRRRRRRRRRR RRTNN  
(R)(20)  
62: RRRRR RRRRRRRRRRRRRRRRRRRR NTNNE  
(R)(20)  
AAM31173 ck: 2324 len: 36 ! Aam31173 Peptide #5210 encoded by probe  
(R,K)(20)  
16: KERKT KRRKKRRRRRRKKRRKKRR R  
(R,K)(20)  
17: ERKTK KRRKKRRRRRRKKRRKKRRR  
(R,K)(20)  
AAM33953 ck: 8343 len: 66 ! Aam33953 Peptide #7990 encoded by probe  
(R,K)(20)  
6: ETERE KKKKKKKKKKKKKKKKKKKKK KKKKK  
(R,K)(20)  
7: TEREK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(R,K)(20)  
8: EREKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(R,K)(20)  
9: REKKK KKKKKKKKKKKKKKKKKKKKK NKKKK  
(R,K)(20)  
AAM36456 ck: 2394 len: 57 ! Aam36456 Peptide #10493 encoded by probe  
(R,K)(20)  
20: EEEEG RRRRRRRRRRRRRRRRRRRR RRRRR  
(R)(20)  
21: EEGRG RRRRRRRRRRRRRRRRRRRR RRRRG  
(R)(20)  
22: EEGRR RRRRRRRRRRRRRRRRRRRR RRRGR  
(R)(20)  
23: EGRRR RRRRRRRRRRRRRRRRRRRR RRRGR  
(R)(20)  
24: GRRRR RRRRRRRRRRRRRRRRRRRR RGRRR  
(R)(20)  
25: RRRRR RRRRRRRRRRRRRRRRRRRR GRRRR  
(R)(20)  
AAM37018 ck: 4228 len: 24 ! Aam37018 Peptide #11055 encoded by probe  
(R,K)(20)  
1: RRRRRRRRRRRRRRRRRRRRRR RKRT  
(R,K)(20)

2: R RRRRRRRRRRRRRRRRRR KRT  
 (R,K){20}  
 3: RR RRRRRRRRRRRRRRRRRR RT  
 (R,K){20}  
 4: RRR RRRRRRRRRRRRRRRRRR T

AA037794 ck: 3301 len: 52 i AA037794 Peptide #11831 encoded by probe fd

12: KKKKN KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 (R,K){20}  
 13: KKKNK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 14: KKNKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 15: KNNKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 16: NKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 17: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 18: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 19: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 20: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 21: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 22: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 23: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 24: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 25: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 26: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 27: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 28: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 29: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 30: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 31: KKKKK KKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}

1

AA038273 ck: 4895 len: 51 i AA038273 Peptide #12310 encoded by probe  
 (R,K){20}  
 18: LFKPM KKKKKKKKKKKKKKKKKR KKLTT  
 (R,K){20}  
 19: FKPMR KKKKKKKKKKKKKKKRRR KLTTT  
 (R,K){20}  
 20: KPMRK RRRKKKKKKKKKKRRRK LTTT  
 (R,K){20}

AA04283 ck: 8137 len: 45 i AA04283 Trimeric fusogenic peptide #2 u

4: YKA KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 (K){20}  
 5: YKA KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 6: YKA KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 7: KAKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 8: AKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 9: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 10: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 11: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 12: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 13: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 14: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 15: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 16: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 17: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 18: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 19: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 20: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 21: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}  
 22: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 (R,K){20}



1

AU04285 ck: 4361 len: 59 i Au04285 Nuclear ligand #2 used in nucleic

23: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKK

24: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK WK

18: AПYKA (R,K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

19: PYKAK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

20: YKAKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

21: KAKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

22: AKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

23: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

24: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

25: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

26: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

27: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

28: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

29: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

30: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

31: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

32: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

33: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

34: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

35: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

36: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

37: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKKK

38: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKK WK

1

AU04287 ck: 4925 len: 100 i Au04287 Poly-L-Lysine used in nucleic

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK

2: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK

3: KK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

9: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

10: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

11: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

12: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

13: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

14: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

15: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

16: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

17: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

18: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

19: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

20: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

21: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

22: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

23: KKKKK (K) [20] KKKKKKKKKKKKKKKKKKKKK KKKKK

```
24: kkkk (K){20} kkkk
25: kkkk (K){20} kkkk
26: kkkk (K){20} kkkk
27: kkkk (K){20} kkkk
28: kkkk (K){20} kkkk
29: kkkk (K){20} kkkk
30: kkkk (K){20} kkkk
31: kkkk (K){20} kkkk
32: kkkk (K){20} kkkk
33: kkkk (K){20} kkkk
34: kkkk (K){20} kkkk
35: kkkk (K){20} kkkk
36: kkkk (K){20} kkkk
37: kkkk (K){20} kkkk
38: kkkk (K){20} kkkk
39: kkkk (K){20} kkkk
40: kkkk (K){20} kkkk
41: kkkk (K){20} kkkk
42: kkkk (K){20} kkkk
43: kkkk (K){20} kkkk
44: kkkk (K){20} kkkk
45: kkkk (K){20} kkkk
46: kkkk (K){20} kkkk
47: kkkk (K){20} kkkk
48: kkkk kkkkkkkkkkkkkkkkk kkkk
49: kkkk (K){20} kkkk
50: kkkk (K){20} kkkk
51: kkkk (K){20} kkkk
52: kkkk (K){20} kkkk
53: kkkk (K){20} kkkk
54: kkkk (K){20} kkkk
55: kkkk (K){20} kkkk
56: kkkk (K){20} kkkk
57: kkkk (K){20} kkkk
58: kkkk (K){20} kkkk
59: kkkk (K){20} kkkk
60: kkkk (K){20} kkkk
61: kkkk (K){20} kkkk
62: kkkk (K){20} kkkk
63: kkkk (K){20} kkkk
64: kkkk (K){20} kkkk
65: kkkk (K){20} kkkk
66: kkkk (K){20} kkkk
67: kkkk (K){20} kkkk
68: kkkk (K){20} kkkk
69: kkkk (K){20} kkkk
70: kkkk (K){20} kkkk
71: kkkk (K){20} kkkk
72: kkkk (K){20} kkkk
```

73: KKKKK (K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
74: KKKKK (K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
75: KKKKK (K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
76: KKKKK (K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
77: KKKKK (K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
78: KKKKK (K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
79: KKKKK (K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
80: KKKKK (K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
81: KKKKK (K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK

AAm01857 ck: 5383 len: 86 ! Aam01857 Peptide #539 encoded by probe for

15: RRRRG (R,K)(20) RRRRG RRRRG RRRRG RRRRG RRRRG  
16: RRRGR (R,K)(20) RRRGR RRRGR RRRGR RRRGR RRRGR  
17: RRGRR (R,K)(20) RRGRR RRGRR RRGRR RRGRR RRGRR  
18: RGRRR (R,K)(20) RGRRR RGRRR RGRRR RGRRR RGRRR  
19: GRRRR (R,K)(20) GRRRR GRRRR GRRRR GRRRR GRRRR  
20: RRRRR (R,K)(20) RRRRR RRRRR RRRRR RRRRR RRRRR  
21: RRRRK (R,K)(20) RRRRK RRRRK RRRRK RRRRK RRRRK  
22: RRRKK (R,K)(20) RRRKK RRRKK RRRKK RRRKK RRRKK  
23: RRKKK (R,K)(20) RRKKK RRKKK RRKKK RRKKK RRKKK  
24: RKKKK (R,K)(20) RKKKK RKKKK RKKKK RKKKK RKKKK  
25: KKKKK (R,K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
26: KKKKK (R,K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
27: KKKKK (R,K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
28: KKKKK (R,K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK

1

29: KKKKK (R,K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
30: KKKKK (R,K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
31: KKKKK (R,K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
32: KKKKK (R,K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
33: KKKKK (R,K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
34: KKKKK (R,K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
35: KKKKK (R,K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
36: KKKKK (R,K)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
37: KKKKK (R)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
38: KKKKK (R)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
39: KKKKK (R)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
40: KKKKK (R)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
41: KKKKK (R)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
42: KKKKK (R)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
43: KKKKK (R)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
44: KKKKK (R)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
45: KKKKK (R)(20) KKKKK KKKKK KKKKK KKKKK KKKKK  
AAm02687 ck: 1334 len: 86 ! Aam02687 Peptide #1369 encoded by probe  
57: EEEEG (R,K)(20) EEEEG EEEEG EEEEG EEEEG EEEEG  
58: EEEGR (R,K)(20) EEEGR EEEGR EEEGR EEEGR EEEGR  
59: EEEGR (R,K)(20) EEEGR EEEGR EEEGR EEEGR EEEGR  
60: EEEGR (R,K)(20) EEEGR EEEGR EEEGR EEEGR EEEGR  
61: GRRRK (R,K)(20) GRRRK GRRRK GRRRK GRRRK GRRRK  
62: RRRKK (R,K)(20) RRRKK RRRKK RRRKK RRRKK RRRKK

(R,K)(20)  
63: RRRKK KRRKKKKKKKKKKKKKKKK KKKK  
(R,K)(20)  
64: RKKKK RRRKKKKKKKKKKKKKKKK KKK  
(R,K)(20)  
65: KKKKR KRRKKKKKKKKKKKKKKKK KK  
(R,K)(20)  
66: KKKRK RKKKKKKKKKKKKKKKKKK K  
(K)(20)  
67: KKKR KKKKKKKKKKKKKKKKK

1  
AAM02768 ck: 9082 len: 167 i Aam02768 Peptide #1450 encoded by probe for  
(R,K)(20)  
33: EGRG RRRRRRRRRRRRRRRRRR RRG  
(R,K)(20)  
34: LGRG RRRRRRRRRRRRRRRRRR RGG  
(R,K)(20)  
35: GGRG RRRRRRRRRRRRRRRRRR GGG  
(R,K)(20)  
AAM03278 ck: 2276 len: 89 i Aam03278 Peptide #1960 encoded by probe for  
(R,K)(20)  
23: EEEE KRRKKKKKKKKKKKKKKKK EEEK  
(R,K)(20)  
46: KEEE KRRKKKKKKKKKKKKKKKK KREE  
(K)(20)  
47: KEEK KRRKKKKKKKKKKKKKKKK KKEE  
(R,K)(20)  
48: EEEK KRRKKKKKKKKKKKKKKKK KEEE  
(R,K)(20)  
49: EEEK KRRKKKKKKKKKKKKKKKK EEEE

1  
AAM03564 ck: 1939 len: 130 i Aam03564 Peptide #2246 encoded by probe for  
(R,K)(20)  
42: EGKE RRRRRRRRRRRRRRRRRR RRRR  
(R,K)(20)  
43: GKER RRRRRRRRRRRRRRRRRR RRRR  
(R,K)(20)  
44: KEER RRRRRRRRRRRRRRRRRR RRRR  
(R,K)(20)  
45: KEER RRRRRRRRRRRRRRRRRR RRRR  
(R,K)(20)  
46: EERR RRRRRRRRRRRRRRRRRR RRRR  
(R,K)(20)  
47: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R,K)(20)  
48: RRRR RRRRRRRRRRRRRRRRRR RRRR

(R,K)(20)  
49: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R,K)(20)  
50: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R,K)(20)  
51: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R,K)(20)  
52: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R,K)(20)  
53: RRRR RRRRRRRRRRRRRRRRRR RRRR  
(R,K)(20)  
54: RRRR RRRRRRRRRRRRRRRRRR RRRR

1  
AAM03848 ck: 3607 len: 88 i Aam03848 Peptide #2530 encoded by probe  
(R,K)(20)  
39: ERKE KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
40: RREK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
41: KREK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
42: REKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
43: ERKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
44: KKKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
45: KKKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
46: KKKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
47: KKKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
48: KKKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
49: KKKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
50: KKKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
51: KKKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
52: KKKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
53: KKKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
54: KKKK KRRKKKKKKKKKKKKKKKK KRRR  
(K)(20)  
55: KKKK KRRKKKKKKKKKKKKKKKK KRRR

```

56:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
57:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
58:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
59:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
60:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
61:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
62:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
63:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
64:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
65:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
66:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKK
67:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KK
68:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK K
69:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK

AAM04408 ck: 3937 len: 85 i Aam04408 Peptide #3090 encoded by probe for
1:  (R,K)(20) KKKKKKKKKKKKKKKKKKK KKKK
2:  K (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
3:  KK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
4:  KKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
5:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
6:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKK
7:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKKE
8:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKKKE
9:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKEEE

```

```

1
AAM06100 ck: 1560 len: 88 i Aam06100 Peptide #4782 encoded by probe
10:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK KKEEE
11:  KKKK (K)(20) KKKKKKKKKKKKKKKKKKK EEEEX
43:  KRRG (R,K)(20) KRRRRRRRRRRRRRRRRRR RRRR
44:  ERGR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
45:  RGRG (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
46:  RGRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
47:  GRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
48:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
49:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
50:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
51:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
52:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
53:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
54:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
55:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
56:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
57:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
58:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
59:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
60:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
61:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
62:  RRRR (R)(20) RRRRRRRRRRRRRRRRRRR RRRR
AAG73687 ck: 3063 len: 29 i Aag73687 Human colon cancer antigen proi

```

1  
(R,K){20}  
8: MMATF KKKKKKKKKKKKKKKKK KX  
(K){20}  
9: MMTEK KKKKKKKKKKKKKKKKK X

AAG73729 ck: 783 len: 83 1 Aag73729 Human colon cancer antigen protein  
(R,K){20}  
49: LGPCE KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
50: GPCEK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
51: PCEKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
52: CEKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
53: EKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKK KKGK  
(K){20}  
55: KKKKK KKKKKKKKKKKKKKKKK KGGK  
(K){20}  
56: KKKKK KKKKKKKKKKKKKKKKK GGRK  
(K){20}  
AAG73810 ck: 374 len: 88 1 Aag73810 Human colon cancer antigen protein  
(R,K){20}  
44: PGOTX KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
45: GQTXK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
46: QTXKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
47: TXKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
48: XKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
50: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
51: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

1  
55: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
56: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
57: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
58: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
62: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
63: KKKKK KKKKKKKKKKKKKKKKK KKGK  
(K){20}  
64: KKKKK KKKKKKKKKKKKKKKKK KGGK  
(K){20}  
65: KKKKK KKKKKKKKKKKKKKKKK GGRK  
(K){20}  
AAG73895 ck: 1887 len: 43 1 Aag73895 Human colon cancer antigen prot  
(R,K){20}  
18: VPRV RKKKKKKKKKKKKKKKK KKKK  
(R,K){20}  
19: RPRV RKKKKKKKKKKKKKKKK KKKK  
(K){20}  
20: PVRK KKKKKKKKKKKKKKKKK KKGK  
(K){20}  
21: RVRK KKKKKKKKKKKKKKKKK KKGK  
(K){20}  
22: VRKK KKKKKKKKKKKKKKKKK GK  
(K){20}  
AAG74218 ck: 8659 len: 104 1 Aag74218 Human colon cancer antigen prot  
(R,K){20}  
75: PLGGQ KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
76: LGGQK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
77: GGQKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
78: GQKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
79: QKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
80: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

```
1
AAG74527 ck: 2664 len: 40 i Aag74527 Human colon cancer antigen protein
(R,K){20}
(K){20}
9: CLLLY KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
10: LLLYK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
11: LLYKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
12: LYKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
13: YKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
GXPX

AAG74650 ck: 1596 len: 69 i Aag74650 Human colon cancer antigen protein
(R,K){20}
(K){20}
36: LQCRQ KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
37: OCRQK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
38: CROKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
39: ROKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
40: OKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
GXPX

AAG74793 ck: 8497 len: 152 i Aag74793 Human colon cancer antigen protein
(R,K){20}
(K){20}
```

```
1
122: SHTQ KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
123: SHTQK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AAG74907 ck: 1215 len: 98 i Aag74907 Human colon cancer antigen pro
(R,K){20}
(K){20}
57: NLBKE KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
58: LRKEK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
59: RLEKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
60: KEKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
61: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
70: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
71: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
72: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
73: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
74: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
75: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
76: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
XGG

AAG75215 ck: 3913 len: 155 i Aag75215 Human colon cancer antigen pro
(R,K){20}
```

(K){20}  
135: RSSAP KKKKKKKKKKKKKKKKK K  
(K){20}  
136: SSAP KKKKKKKKKKKKKKKKK

1  
AAG75886 ck: 4235 len: 71 i Aag75886 Human colon cancer antigen protein  
(R,K){20}  
(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKK KXGX  
(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKK XGX

1  
AAE01796 ck: 4416 len: 72 i Aae01796 Human gene 27 encoded secreted pro  
(R,K){20}  
(K){20}  
47: LPTFL KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: PTFLK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: TPLKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
50: FLKKK KKKKKKKKKKKKKKKKK KKK  
(K){20}  
51: LKKKK KKKKKKKKKKKKKKKKK KK  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKK K  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKK

1  
AAE01848 ck: 5584 len: 73 i Aae01848 Human gene 27 encoded secreted pro  
(R,K){20}  
(K){20}  
47: LPTFL KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: PTFLK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: TPLKK KKKKKKKKKKKKKKKKK KKKKI  
(K){20}  
50: FLKKK KKKKKKKKKKKKKKKKK KKKI  
(K){20}  
51: LKKKK KKKKKKKKKKKKKKKKK KKI  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKK KI  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKK I  
AAB90574 ck: 1431 len: 530 i Aab90574 Human secreted protein, SEQ ID NO:  
(R,K){20}  
(K){20}  
511: LHAPP KKKKKKKKKKKKKKKKK

1  
AAB45846 ck: 8137 len: 45 i Aab45846 Nucleic acid transporter system  
(R,K){20}  
(K){20}  
4: YKA KKKKKKKKKKKKKKKKK KKKKK

5: YKAK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: YKAKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KAKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

8: AKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKK KKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKK KK

1  
AAB45848 ck: 4361 len: 59 i Aab45848 Nucleic acid transporter system  
(R,K){20}  
(K){20}  
18: APYKA KKKKKKKKKKKKKKKKK KKKKK



```

19: РYAKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
20: YKAKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
21: KAKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
22: AKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
23: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
24: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
25: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
26: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
27: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
28: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
29: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
30: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
31: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
32: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
33: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
34: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
35: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
36: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
37: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201
38: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K)1201

AAB45850 ck: 4925 len: 100 1 Aab45850 Nucleic acid transporter system pe
(R,K)1201
(K)1201
KKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
1:
(K)1201
2: K KKKKKKKKKKKKKKKKKKKKK KKKKK
(K)1201
3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K)1201

```

[illegible]

28: kkkk (K){20} kkkk  
29: kkkk (K){20} kkkk  
30: kkkk (K){20} kkkk  
31: kkkk (K){20} kkkk  
32: kkkk (K){20} kkkk  
33: kkkk (K){20} kkkk  
34: kkkk (K){20} kkkk  
35: kkkk (K){20} kkkk  
36: kkkk (K){20} kkkk  
37: kkkk (K){20} kkkk  
38: kkkk (K){20} kkkk  
39: kkkk (K){20} kkkk  
40: kkkk (K){20} kkkk  
41: kkkk (K){20} kkkk  
42: kkkk (K){20} kkkk  
43: kkkk (K){20} kkkk  
44: kkkk (K){20} kkkk  
45: kkkk (K){20} kkkk  
46: kkkk (K){20} kkkk  
47: kkkk (K){20} kkkk  
48: kkkk (K){20} kkkk  
49: kkkk (K){20} kkkk  
50: kkkk (K){20} kkkk  
51: kkkk (K){20} kkkk

---

52: kkkk kkkkkkkkkkkkkkkkk kkkk  
53: kkkk (K){20} kkkk  
54: kkkk (K){20} kkkk  
55: kkkk (K){20} kkkk  
56: kkkk (K){20} kkkk  
57: kkkk (K){20} kkkk  
58: kkkk (K){20} kkkk  
59: kkkk (K){20} kkkk  
60: kkkk (K){20} kkkk  
61: kkkk (K){20} kkkk  
62: kkkk (K){20} kkkk  
63: kkkk (K){20} kkkk  
64: kkkk (K){20} kkkk  
65: kkkk (K){20} kkkk  
66: kkkk (K){20} kkkk  
67: kkkk (K){20} kkkk  
68: kkkk (K){20} kkkk  
69: kkkk (K){20} kkkk  
70: kkkk (K){20} kkkk  
71: kkkk (K){20} kkkk  
72: kkkk (K){20} kkkk  
73: kkkk (K){20} kkkk  
74: kkkk (K){20} kkkk  
75: kkkk (K){20} kkkk  
76: kkkk (K){20} kkkk

77: KKKK (K){20} KKKK  
 78: KKKK (K){20} KKKK  
 79: KKKK (K){20} KKKK  
 80: KKKK (K){20} KKKK  
 81: KKKK (K){20} KKKK

AAB50247 ck: 8085 len: 154 ! Aab50247 Human breast cancer associated B72

114: TOLRO (R,K){20} KKKK  
 115: QLPQK (K){20} KKKK  
 116: LPQK (K){20} KKKK  
 117: RQK (K){20} KKKK  
 118: QK (K){20} KKKK  
 119: KKKK (K){20} KKKK  
 120: KKKK (K){20} KKKK  
 121: KKKK (K){20} KKKK  
 122: KKKK (K){20} KKKK  
 123: KKKK (K){20} KKKK  
 124: KKKK (K){20} KKKK  
 125: KKKK (K){20} KKKK  
 126: KKKK (K){20} KKKK  
 127: KKKK (K){20} KKKK  
 128: KKKK (K){20} KKKK

ABB44830 ck: 5275 len: 38 ! ABB44830 Human protective sequence CNI-0074

4: MAQ (R,K){20} KKKK  
 (K){20}

5: MAQ KKKK (K){20} KKKK  
 6: MAQK (K){20} KKKK  
 7: AQK (K){20} KKKK  
 8: QK (K){20} KKKK  
 9: KKKK (K){20} KKKK  
 10: KKKK (K){20} KKKK  
 11: KKKK (K){20} KKKK  
 12: KKKK (K){20} KKKK  
 13: KKKK (K){20} KKKK  
 14: KKKK (K){20} KKKK  
 15: KKKK (K){20} KKKK  
 16: KKKK (K){20} KKKK  
 17: KKKK (K){20} KKKK

AAU69690 ck: 875 len: 46 ! AAU69690 Cell death protective sequence

15: RKKK (R,K){20} KKKK  
 16: EKSK (K){20} KKKK  
 17: KSKK (K){20} KKKK  
 18: KSKK (K){20} KKKK  
 19: SKKK (K){20} KKKK  
 20: KKKK (K){20} KKKK  
 21: KKKK (K){20} KKKK  
 22: KKKK (K){20} KKKK  
 23: KKKK (K){20} KKKK  
 24: KKKK (K){20} KKKK

25: KKKKK KKKKKKKKKKKKKKKKKKKKK GG

1

AAU69736 ck: 5862 len: 50 i Aau69736 Cell death protective sequence CN1

(R,K){20}  
(R,K){20}

11: ESALG RKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

12: SALGR KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

13: ALGRK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

14: LGRKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

15: GRKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

16: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

23: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

25: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

26: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

27: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

28: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

29: KKKKK KKKKKKKKKKKKKKKKKKKKK GR

Databases searched:

EMBL, Release 7.0, Released on 28Apr2002, Formatted on 2May2002

Total finds: 7,093

Total length: 114,001,827

CPU time: 12:26.84